Communications in Computer and Information Science

716

Commenced Publication in 2007 Founding and Former Series Editors: Alfredo Cuzzocrea, Dominik Ślęzak, and Xiaokang Yang

Editorial Board

Simone Diniz Junqueira Barbosa

Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Rio de Janeiro. Brazil

Phoebe Chen

La Trobe University, Melbourne, Australia

Xiaoyong Du

Renmin University of China, Beijing, China

Joaquim Filipe

Polytechnic Institute of Setúbal, Setúbal, Portugal

Orhun Kara

TÜBİTAK BİLGEM and Middle East Technical University, Ankara, Turkey

Igor Kotenko

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

Ting Liu

Harbin Institute of Technology (HIT), Harbin, China

Krishna M. Sivalingam

Indian Institute of Technology Madras, Chennai, India

Takashi Washio

Osaka University, Osaka, Japan

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

Stanisław Kozielski · Dariusz Mrozek Paweł Kasprowski · Bożena Małysiak-Mrozek Daniel Kostrzewa (Eds.)

Beyond Databases, Architectures and Structures

Towards Efficient Solutions for Data Analysis and Knowledge Representation

13th International Conference, BDAS 2017 Ustroń, Poland, May 30 – June 2, 2017 Proceedings



Editors
Stanisław Kozielski
Institute of Informatics
Silesian University of Technology
Gliwice
Poland

Dariusz Mrozek

Institute of Informatics

Silesian University of Technology

Gliwice

Poland

Paweł Kasprowski

Institute of Informatics

Silesian University of Technology

Gliwice

Poland

Bożena Małysiak-Mrozek Institute of Informatics Silesian University of Technology Gliwice Poland

Daniel Kostrzewa
Institute of Informatics
Silesian University of Technology
Gliwice
Poland

ISSN 1865-0929 ISSN 1865-0937 (electronic) Communications in Computer and Information Science ISBN 978-3-319-58273-3 ISBN 978-3-319-58274-0 (eBook) DOI 10.1007/978-3-319-58274-0

Library of Congress Control Number: 2017938564

© Springer International Publishing AG 2017

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

Collecting, processing, and analyzing data have become important branches of computer science. Many areas of our existence generate a wealth of information that must be stored in a structured manner and processed appropriately in order to gain the knowledge from the inside. Databases have become a ubiquitous way of collecting and storing data. They are used to hold data describing many areas of human life and activity, and as a consequence, they are also present in almost every IT system. Today's databases have to face the problem of data proliferation and growing variety. More efficient methods for data processing are needed more than ever. New areas of interests that deliver data require innovative algorithms for data analysis.

Beyond Databases, Architectures and Structures (BDAS) is a series of conferences located in Central Europe and very important for this geographic region. The conference intends to give the state of the art of the research that satisfies the needs of modern, widely understood database systems, architectures, models, structures, and algorithms focused on processing various types of data. The aim of the conference is to reflect the most recent developments of databases and allied techniques used for solving problems in a variety of areas related to database systems, or even go one step forward — beyond the horizon of existing databases, architectures, and data structures.

The 13th International BDAS Scientific Conference (BDAS 2017), held in Ustroń, Poland, from May 30 to June 2, 2017, was a continuation of the highly successful BDAS conference series started in 2005. For many years BDAS has been attracting hundreds or even thousands of researchers and professionals working in the field of databases. Among attendees of our conference were scientists and representatives of IT companies. Several editions of BDAS were supported by our commercial, world-renowned partners, developing solutions for the database domain, such as IBM, Microsoft, Sybase, Oracle, and others. BDAS annual meetings have become an arena for exchanging information on the widely understood database systems and data-processing algorithms.

BDAS 2017 was the 13th edition of the conference, organized under the technical co-sponsorship of the IEEE Poland Section. We also continued our successful cooperation with Springer, which resulted in the publication of this book. The conference attracted more than a hundred participants from 15 countries, who made this conference a successful and memorable event. There were three keynote talks and one tutorial given by leading scientists: Prof. Jens Allmer from the Department of Molecular Biology and Genetics, Izmir Institute of Technology, Urla, Izmir, gave an excellent keynote talk entitled "Database Integration Facilitating the Merging of MicorRNA and Gene Regulatory Pathways in ALS." Prof. Dirk Labudde from the Bioinformatics group Mittweida (bigM) and Forensic Science Investigation Lab (FoSIL), University of Applied Sciences, Mittweida, Germany, honored us with a presentation entitled "3D Crime Scene and Disaster Site Reconstruction using Open Source Software." Dr. Dominik Szczerba from Future Processing, Gliwice, Poland, gave a talk on "Computational Physiology." Prof. Jean-Charles Lamirel from SYNALP team, LORIA,

Vandœuvre-lès-Nancy, France, prepared a tutorial on "Text Mining in the Big Data Context: Existing Approaches and Challenges." The keynote speeches, tutorials, and plenary sessions allowed participants to gain insight into new areas of data analysis and data processing.

BDAS is focused on all aspects of databases. It is intended to have a broad scope, including different kinds of data acquisition, processing, and storing, and this book reflects fairly well the large span of research presented at BDAS 2017. This volume consists of 44 carefully selected papers that are assigned to seven thematic groups:

- Big data and cloud computing
- Artificial intelligence, data mining, and knowledge discovery
- Architectures, structures, and algorithms for efficient data processing
- Text mining, natural language processing, ontologies, and Semantic Web
- Bioinformatics and biological data analysis
- Industrial applications
- Data mining tools, optimization, and compression

The first group, containing four papers, is devoted to big data and cloud computing. Papers in this group discuss hot topics of stream processing with MapReduce, a tensor-based approach to temporal features modeling with application in big data, querying XML documents with SparkSQL, and automatic scaling computing infrastructure of the cloud. The second group contains six papers devoted to various methods used in data mining, knowledge discovery, and knowledge representation. Papers assembled in this group show a wide spectrum of applications of various exploration techniques, including decision rules, knowledge-based systems, clustering and classification algorithms, and rough sets, to solve many real-world problems.

The third group contains nine papers devoted to various database architectures and models, data structures, and algorithms used for efficient data processing. Papers in this group discuss the effectiveness of query execution, performance, and consistency of various database systems, including relational and NoSQL databases, indexing structures, sorting algorithms, and distributed data processing. The fourth group consists of nine papers devoted to natural language processing, text mining, ontologies, and the Semantic Web. These papers discuss problems of building recommendation systems with the use ontologies, extending expressiveness of knowledge description, ontology reuse for fast prototyping of new concepts, processing natural language instructions by robots, data integration in NLP, authorship attribution for texts, plagiarism detection, and RDF validation.

The research devoted to bioinformatics and biological data analysis is presented in six papers gathered in the fifth group. The papers cover problems connected with gene expression and chromatography but also medical diagnosing as well as face and emotion recognition. The sixth group includes four papers describing various applications of data mining — especially in coal mining and automotive industries. The last group includes six papers presenting various data-mining tools, performance optimization techniques, and a compression algorithm.

We hope that the broad scope of topics related to databases covered in this proceedings volume will help the reader to understand that databases have become an important element of nearly every branch of computer science.

We would like to thank all Program Committee members and additional reviewers for their effort in reviewing the papers. Special thanks to Piotr Kuźniacki — builder and for 12 years administrator of our website bdas.polsl.pl. The conference organization would not have been possible without the technical staff: Dorota Huget and Jacek Pietraszuk.

March 2017

Stanisław Kozielski Dariusz Mrozek Pawel Kasprowski Bożena Małysiak-Mrozek Daniel Kostrzewa

Organization

Program Committee

Chair

Stanisław Kozielski Silesian University of Technology, Poland

Honorary Member

Lotfi A. Zadeh University of California, Berkeley, USA

Members

Alla Anohina-Naumeca Riga Technical University, Latvia Sansanee Auephanwiriyakul Chiang Mai University, Thailand

Vasile Avram Bucharest Academy of Economic Studies, Romania

Sergii Babichev J.E. Purkyně University in Ústí nad Labem,

Czech Republic

Werner Backes Sirrix AG Security Technologies, Bochum, Germany Susmit Bagchi Gyeongsang National University, Jinju, South Korea

Péter Balázs University of Szeged, Hungary

Katalin Balla Budapest University of Technology and Economics,

Hungary

Igor BernikNove Mesto University, SloveniaBora BimbariUniversity of Tirana, AlbaniaMarko BohanecUniversity of Nova Gorica, Slovenia

Alexandru Boicea Polytechnic University of Bucharest, Romania

Patrick Bours Gjovik University College, Norway
Lars Braubach University of Hamburg, Germany
Ljiljana Brkić University of Zagreb, Croatia
Marija Brkić Bakarić University of Rijeka, Croatia

Germanas Budnikas Kaunas University of Technology, Lithuania Peter Butka Technical University of Košice, Slovakia Rita Butkienė Kaunas University of Technology, Lithuania

Sanja Čandrlić University of Rijeka, Croatia

George D.C. Cavalcanti Universidade Federal de Pernambuco, Brazil Chantana Chantrapornchai Kasetsart University, Bangkok, Thailand

Ming Chen University of Bielefeld, Germany

Andrzej Chydziński Silesian University of Technology, Poland

Armin B. Cremers University of Bonn, Germany

Tadeusz Czachórski IITiS, Polish Academy of Sciences, Poland

Yixiang Chen East China Normal University, Shanghai, P.R. China

Po-Yuan Chen China Medical University, Taichung, Taiwan,

University of British Columbia, BC, Canada

Robertas Damaševičius Kaunas University of Technology, Lithuania Mirela Danubianu Ștefan cel Mare University of Suceava, Romania Goce Delčev University of Štip, Macedonia Blagoj Delipetrev Sebastian Deorowicz Silesian University of Technology, Poland József Dombi University of Szeged, Hungary

Jack Dongarra University of Tennessee, Knoxville, USA

Libor Dostálek University of South Bohemia in Ceske Budejovice.

Czech Republic

Ecole Polytechnique Federale de Lausanne, Andrzej Drygajlo

Switzerland

Denis Enachescu University of Bucharest, Romania Zoe Falomir Llansola University of Bremen, Germany

Alexandru Ioan Cuza University, Romania Victor Felea German University of Technology, Oman Rudolf Fleischer Iwate Prefectural University, Japan Hamido Fujita

Rezekne Higher Education Institution, Latvia Peter Grabusts

York University, Ontario, Canada Jarek Gryz Donatella Gubiani University of Nova Gorica, Slovenia Saulius Gudas Vilnius University, Lithuania University of Oradea, Romania Cornelia Gvorodi Robert Gyorodi University of Oradea, Romania

Abdelkader Hameurlain Paul Sabatier University, Toulouse Cedex, France

Cambridge Computed Imaging Ltd., UK Mike Havball Brahim Hnich Izmir University of Economics, Turkey

Saarland University, Germany Jörg Hoffmann

Edward Hrynkiewicz Silesian University of Technology, Poland,

IEEE member

Drexel University, Philadelphia, USA Xiaohua Tony Hu

Jiewen Huang Google, USA

Wroclaw University of Technology, Poland Zbigniew Huzar

Csanád Imreh University of Szeged, Hungary University of Rijeka, Croatia Marina Ivašič-Kos

Karel Ježek University of West Bohemia, Czech Republic

University Džemal Bijedić of Mostar, Emina Junuz

Bosnia and Herzegovina

Silesian University of Technology, Poland, Pawel Kasprowski

IEEE member

Zoltan Kazi University of Novi Sad, Serbia

> Wroclaw University of Technology, Poland IITiS, Polish Academy of Sciences, Poland

Eötvös Loránd University, Hungary Saarland University, Germany

Brno University of Technology, Czech Republic

Silesian University of Technology, Poland

University of Miskolc, Hungary University of Bonn, Germany

Przemysław Kazienko

Jerzy Klamka Attila Kiss

Matthias Klusch

Dušan Kolář Daniel Kostrzewa

László Kovács Stefan Kratsch Tomas Krilavičus Vytautas Magnus University, Lithuania Antonín Kučera Masaryk University, Czech Republic Bora I. Kumova Izmir Institute of Technology, Turkey Andrzej Kwiecień Silesian University of Technology, Poland

Dirk Labudde University of Applied Sciences, Mittweida, Germany Jean-Charles Lamirel LORIA, Nancy, France, University of Strasbourg,

France

Dejan Lavbič University of Ljubljana, Slovenia Fotios Liarokapis Masaryk University, Czech Republic

Sergio Lifschitz Pontificia Universidade Catolica do Rio de Janeiro,

Brazil

Antoni Ligęza AGH University of Science and Technology, Poland

Maciej Liśkiewicz University of Lübeck, Germany

Ivica Lukić Josip Juraj Strossmayer University of Osijek, Croatia

Ivan Luković University of Novi Sad, Serbia

Bożena Małysiak-Mrozek Silesian University of Technology, Poland,

IEEE member

Algirdas Maknickas Vilnius Gediminas Technical University, Lithuania Violeta Manevska St. Clement of Ohrid University of Bitola, Macedonia

Saulius Maskeliūnas Vilnius University, Lithuania

Jelena Mamčenko Vilnius Gediminas Technical University, Lithuania

Marco Masseroli Politecnico di Milano, Italy

Dalius Mažeika Vilnius Gediminas Technical University, Lithuania Zygmunt Mazur Wroclaw University of Technology, Poland Peter Mikulecký University of Hradec Králové, Czech Republic

Biljana Mileva Boshkoska Nove Mesto University, Slovenia Guido Moerkotte University of Mannheim, Germany

Yasser F.O. Mohammad Assiut University, Egypt

Tadeusz Morzy Poznan University of Technology, Poland

Mikhail Moshkov King Abdullah University of Science and Technology,

Saudi Arabia

Dariusz Mrozek Silesian University of Technology, Poland,

IEEE member

Mieczysław Muraszkiewicz Warsaw University of Technology, Poland Sergio Nesmachnow Universidad de la Republica, Uruguay

Laila Niedrīte University of Latvia, Latvia Mladen Nikolić University of Belgrade, Serbia

Tadeusz Pankowski Poznan University of Technology, Poland Martynas Patašius Kaunas University of Technology, Lithuania Bogdan Pătruț Vasile Alecsandri University of Bacău, Romania

Mile Pavlić University of Rijeka, Croatia

Witold Pedrycz
University of Alberta, Canada, IEEE member
Adam Pelikant
Lodz University of Technology, Poland
Horia F. Pop
Babeş-Bolyai University, Romania
Václav Přenosil
Masaryk University, Czech Republic
Hugo Proenca
University of Beira Interior, Portugal

Vvtenis Punvs Kaunas University of Technology, Lithuania

Abdur Rakib University of Nottingham, Semenvih, Selangor D.E.

Malaysia

University of North Carolina, Charlotte, USA Zbigniew W. Ras

Riccardo Rasconi Italian National Research Council, Italy

Jan Rauch University of Economics, Prague, Czech Republic Marek Reiman-Greene

Centre for Applied Science and Technology in Home

Office Science, UK

Silesian University of Technology, Poland Jerzy Respondek

St. Clement of Ohrid University of Bitola, Macedonia Blagoj Ristevski

Blaž Rodič Nove Mesto University, Slovenia

Henryk Rybiński Warsaw University of Technology, Poland Christoph Schommer University of Luxembourg, Luxembourg Roman Šenkeřík Tomas Bata University in Zlín, Czech Republic Rzeszow University of Technology, Poland Galina Setlak

Silesian University of Technology and EMAG, Poland Marek Sikora

Krzysztof Stencel University of Warsaw, Poland

Maria Curie-Skłodowska University, Poland Przemysław Stpiczyński

Dan Mircea Suciu Babes-Bolyai University, Romania

Dominik Ślezak University of Warsaw, Poland, Infobright Inc., Canada

Andrzej Świerniak Silesian University of Technology, Poland Jüri Vain Tallinn University of Technology, Estonia

Czech Technical University in Prague, Czech Republic Michal Valenta

Agnes Vathy-Fogarassy University of Pannonia, Hungary University of Melbourne, Australia Karin Verspoor

Tallinn University, Estonia Sirie Virkus Boris Vrdoljak University of Zagreb, Croatia

Intel Corporation, San Diego, USA, IEEE member Sylwester Warecki

Tadeusz Wieczorek Silesian University of Technology, Poland

University of Göttingen, Germany Lena Wiese

Robert Wrembel Poznan University of Technology, IEEE member,

Poland

Moawia Elfaki Yahia Eldow King Faisal University, Saudi Arabia IITiS, Polish Academy of Sciences, Poland Mirosław Zaborowski

Grzegorz Zareba University of Arizona, Tucson, USA

František Zbořil Brno University of Technology, Czech Republic Jaroslav Zendulka Brno University of Technology, Czech Republic

Iveta Zolotová Technical University of Košice, Slovakia

Xiamen University, P.R. China Ouan Zou University of Latvia, Latvia Jānis Zuters

Organizing Committee

Bożena Małysiak-Mrozek Daniel Kostrzewa Dariusz Mrozek Piotr Kuźniacki Pawel Kasprowski Dorota Huget

Additional Reviewers

Dariusz Rafał Augustyn Thi Hoa Hue Nguyen

Małgorzata Bach Agnieszka Nowak-Brzezińska

Piotr Bajerski Karolina Nurzyńska Robert Brzeski Ewa Płuciennik Adam Duszenko Arkadiusz Poteralski

Jacek Fraczek Ewa Romuk

Katarzyna Harężlak Robert Tutajewicz Michał Kozielski Aleksandra Werner Marcin Michalak Łukasz Wyciślik Alina Momot Hafed Zghidi

Sponsoring Institutions

Technical co-sponsorship of the IEEE Poland Section

Contents

Big Data and Cloud Computing	
Integrating Map-Reduce and Stream-Processing for Efficiency (MRSP) Pedro Martins, Maryam Abbasi, José Cecílio, and Pedro Furtado	3
Tensor-Based Modeling of Temporal Features for Big Data CTR Estimation	16
Evaluation of XPath Queries Over XML Documents Using SparkSQL Framework	28
Metrics-Based Auto Scaling Module for Amazon Web Services Cloud Platform	42
Artificial Intelligence, Data Mining and Knowledge Discovery	
Comparison of Two Versions of Formalization Method for Text Expressed Knowledge	55
Influence of Similarity Measures for Rules and Clusters on the Efficiency of Knowledge Mining in Rule-Based Knowledge Bases	67
Attribute Reduction in a Dispersed Decision-Making System with Negotiations	79
Adjusting Parameters of the Classifiers in Multiclass Classification	89
Data Mining - A Tool for Migration Stock Prediction	102
A Survey on Data Mining Methods for Clustering Complex Spatiotemporal Data	115

Architectures, Structures and Algorithms for Efficient Data Processing	
Multi-partition Distributed Transactions over Cassandra-Like Database with Tunable Contention Control	129
The Multi-model Databases – A Review	141
Comparative Analysis of Relational and Non-relational Databases in the Context of Performance in Web Applications	153
Using Genetic Algorithms to Optimize Redundant Data Iwona Szulc, Krzysztof Stencel, and Piotr Wiśniewski	165
Interoperable SQLite for a Bare PC	177
FM-index for Dummies	189
Lattice Based Consistent Slicer and Topological Cut for Distributed Computation in Monotone Spaces	202
Storage Efficiency of LOB Structures for Free RDBMSs on Example of PostgreSQL and Oracle Platforms	212
Optimization of Memory Operations in Generalized Search Trees	
of PostgreSQL	224
Text Mining, Natural Language Processing, Ontologies and Semantic Web	
Sorting Data on Ultra-Large Scale with RADULS: New Incarnation of Radix Sort	235
Serendipitous Recommendations Through Ontology-Based Contextual Pre-filtering	246
Extending Expressiveness of Knowledge Description with Contextual Approach	260

Contents	XVII
Ontology Reuse as a Means for Fast Prototyping of New Concepts Igor Postanogov and Tomasz Jastrząb	273
Reading Comprehension of Natural Language Instructions by Robots Irena Markievicz, Minija Tamosiunaite, Daiva Vitkute-Adzgauskiene, Jurgita Kapociute-Dzikiene, Rita Valteryte, and Tomas Krilavicius	288
A New Method of XML-Based Wordnets' Data Integration	302
Authorship Attribution for Polish Texts Based on Part of Speech Tagging <i>Piotr Szwed</i>	316
Fast Plagiarism Detection in Large-Scale Data	329
RDF Validation: A Brief Survey	344
Bioinformatics and Biological Data Analysis	
Objective Clustering Inductive Technology of Gene Expression Sequences Features	359
Novel Computational Techniques for Thin-Layer Chromatography (TLC) Profiling and TLC Profile Similarity Scoring	373
Extending the Doctrine ORM Framework Towards Fuzzy Processing of Data: Exemplified by Ambulatory Data Analysis	386
Segmenting Lungs from Whole-Body CT Scans	403
Improved Automatic Face Segmentation and Recognition for Applications with Limited Training Data	415
Emotion Recognition: The Influence of Texture's Descriptors on Classification Accuracy	427

XVIII Contents

Industrial Applications

The Use of the TGSP Module as a Database to Identify Breaks in the Work of Mining Machinery	441
A Data Warehouse as an Indispensable Tool to Determine the Effectiveness of the Use of the Longwall Shearer	453
Computer Software Supporting Rock Stress State Assessment for Deep Coal Mines	466
An Ontology Model for Communicating with an Autonomous Mobile Platform	480
Data Mining Tools, Optimization and Compression	
Relational Transition System in Maude	497
A Performance Study of Two Inference Algorithms for a Distributed Expert System Shell	512
DUABI - Business Intelligence Architecture for Dual Perspective Analytics Bartosz Czajkowski and Teresa Zawadzka	527
Comparative Analysis of JavaScript and Its Extensions for Web Application Optimization	539
ALMM Solver - Database Structure and Data Access Layer Architecture Krzysztof Rączka and Edyta Kucharska	551
Human Visual System Inspired Color Space Transform in Lossy JPEG 2000 and JPEG XR Compression	564
Author Index	577