Lecture Notes in Computer Science **11708**

Founding Editors

Gerhard Goos
Karlsruhe Institute of Technology, Karlsruhe, Germany
Juris Hartmanis
Cornell University, Ithaca, NY, USA

Editorial Board Members

Elisa Bertino
Purdue University, West Lafayette, IN, USA
Wen Gao
Peking University, Beijing, China
Bernhard Steffen
TU Dortmund University, Dortmund, Germany
Gerhard Woeginger
RWTH Aachen, Aachen, Germany
Moti Yung
Columbia University, New York, NY, USA

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

Carlos Ordonez · Il-Yeol Song ·
Gabriele Anderst-Kotsis ·
A Min Tjoa · Ismail Khalil (Eds.)

Big Data Analytics and Knowledge Discovery

21st International Conference, DaWaK 2019
Linz, Austria, August 26–29, 2019
Proceedings

Springer

Editors
Carlos Ordonez
University of Houston
Houston, TX, USA

Gabriele Anderst-Kotsis
Johannes Kepler University of Linz
Linz, Austria

Ismail Khalil
Johannes Kepler University of Linz
Linz, Austria

Il-Yeol Song
Drexel University
Philadelphia, PA, USA

A Min Tjoa
Software Competence Center Hagenberg
Hagenberg im Mühlkreis, Austria

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-030-27519-8 ISBN 978-3-030-27520-4 (eBook)
https://doi.org/10.1007/978-3-030-27520-4

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

© Springer Nature Switzerland AG 2019
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

Big Data Analytics and Knowledge Discovery remain hot research areas for both academia and the software industry, further fueled by advances in hardware and software. Important research topics associated to these major themes include data lakes (schema-free repositories), database design (ER modeling, prototyping), data integration (especially linking structured and semistructured data sources), big data management (mixing relational tables, text and any files), query languages (SQL and beyond), scalable analytic algorithms, parallel systems (cloud, parallel database systems, Spark, MapReduce, HDFS), theoretical foundations, and practical applications.

With a track record of 21 editions, the International Conference on Big Data Analytics and Knowledge Discovery (DaWaK) has established itself as a high-quality forum for researchers, practitioners, and developers in the field of Big Data Analytics. This year's conference (DaWaK 2019) builds on this tradition, facilitating the interdisciplinary exchange of ideas, theory, techniques, experiences, and future research directions. DaWaK 2019 aims to introduce innovative principles, methods, models, algorithms, industrial products, and experiences to solve challenging problems faced in the development of new generation data management and analytic systems in the Big Data era.

Our call for papers attracted 61 submissions, from which the Program Committee finally selected 22 papers, yielding an acceptance rate of 36%. Each paper was reviewed by an average of four reviewers and in some cases up to five. Accepted papers cover a number of broad research areas on both theoretical and practical aspects. Some trends found in accepted papers include new generations of data warehouses, data lakes, data pre-processing, data mining, cloud computing, query processing, sequences, graph analytics, privacy-preserving data mining, and parallel processing. On the other hand, the program featured interesting case-studies on social networks, Twitter sentiment analysis, understanding ground transportation modes, and E-commerce, among others.

Due to the history and reputation of DaWaK, editors of a well-known journal agreed to receive extended versions of best papers selected from our program. This year, we are pleased to have a special issue in: *Data and Knowledge Engineering* (DKE, Elsevier).

We would like to thank all authors for submitting their papers to DaWaK 2019 and we hope they submit again in the future. On the other hand, we express our gratitude to all the Program Committee members who provided high quality reviews. We appreciate the great efforts of Amin Anjomshoaa for helping extend the ConfDriver system with several innovations to improve paper reviews, to manage a conference-to-journal long-term review process. Finally, we would like to thank the DEXA conference organizers for the support and guidance. For conference attendants, we hope they enjoyed the technical program, informal meetings, and interaction with

colleagues from all over the world. For the readers of these proceedings, we hope these papers are interesting and they give you ideas for future research.

August 2019

Carlos Ordonez
Il-Yeol Song

Organization

General Chair

Thomas Natschläger	Software Competence Center Hagenberg, Austria

Program Committee Chairs

Carlos Ordonez	University of Houston, USA
Il-Yeol Song	Drexel University, USA

Steering Committee

Gabriele Anderst-Kotsis	Johannes Kepler University Linz, Austria
Ismail Khalil	Johannes Kepler University Linz, Austria
A Min Tjoa	TU Wien, Austria

Program Committee and Reviewers

Alberto Abellö	Universitat Politecnica de Catalunya, Spain
Toshiyuki Amagasa	University of Tsukuba, Japan
Elena Baralis	Politecnico di Torino, Italy
Ladjel Bellatreche	ENSMA, France
Sadok. Ben Yahia	Faculty of Sciences of Tunis, Tunisia
Jorge Bernardino	ISEC, Polytechnic Institute of Coimbra, Portugal
Vasudha Bhatnagar	Delhi University, India
Omar Boussaid	University of Lyon/Lyon 2, France
Michael Brenner	University of Hannover, Germany
Stephane Bressan	National University of Singapore, Singapore
Wellington Cabrera	Teradata, USA
Joel Luìs Carbonera	Federal University of Rio Grande do Sul, Brazil
Sharma Chakravarthy	The University of Texas at Arlington, USA
Frans Coenen	The University of Liverpool, UK
Isabelle Comyn-Wattiau	ESSEC Business School Paris, France
Alfredo Cuzzocrea	University of Trieste, Italy
Laurent d'Orazio	University of Rennes 1, France
Soumyava Das	Teradata, USA
Karen Davis	Miami University, USA
Claudia Diamantini	Università Politecnica delle Marche, Italy
Josep Domingo-Ferrer	Universitat Rovira i Virgili, Spain
Dejing Dou	University of Oregon, USA
Markus Endres	University of Augsburg, Germany
Leonidas Fegaras	The University of Texas at Arlington, USA

Philippe Fournier-Viger	Harbin Institute of Technology, China
Filippo Furfaro	Universitat della Calabria, Italy
Pedro Furtado	Universidade de Coimbra, Portugal
Carlos Garcia-Alvarado	Autonomic LLC, USA
Kazuo Goda	The University of Tokyo, Japan
Matteo Golfarelli	DISI - University of Bologna, Italy
Sergio Greco	University of Calabria, Italy
Hyoil Han	Illinois State University, USA
Takahiro Hara	Osaka University, Japan
Frank Hoppner	Ostfalia University of Applied Sciences, Germany
Stéphane Jean	LIAS/ISAE-ENSMA, University of Poitiers, France
Selma Khouri	LCSI/ESI, Algeria, and LIAS/ISAE-ENSMA, France
Min-Soo Kim	DGIST, Republic of Korea
Uday Kiran Rage	University of Tokyo, Japan
Jens Lechtenboerger	Westfalische Wilhelms-Universitat Munster, Germany
Young-Koo Lee	Kyung Hee University, Republic of Korea
Jae-Gil Lee	KAIST, Republic of Korea
Carson Leung	University of Manitoba, Canada
Sebastian Link	The University of Auckland, New Zealand
Sofian Maabout	University of Bordeaux, France
Patrick Marcel	Universite Francois Rabelais Tours, France
Alejandro Mate	University of Alicante, Spain
Jun Miyazaki	Tokyo Institute of Technology, Japan
Anirban Mondal	Ashoka University, India
Yang-Sae Moon	Kangwon National University, Republic of Korea
Yasuhiko Morimoto	Hiroshima University, Japan
Makoto Onizuka	Osaka University, Japan
Alex Poulovassilis	Birkbeck, University of London, UK
Praveen Rao	University of Missouri-Kansas City, USA
Franck Ravat	IRIT, Université Toulouse I Capitole, France
Goce Ristanoski	Officeworks, Australia

Organizers

Institute for
Telecooperation

Contents

Applications

Detecting the Onset of Machine Failure Using Anomaly Detection Methods 3
Mohammad Riazi, Osmar Zaiane, Tomoharu Takeuchi, Anthony Maltais, Johannes Günther, and Micheal Lipsett

A Hybrid Architecture for Tactical and Strategic Precision Agriculture 13
Enrico Gallinucci, Matteo Golfarelli, and Stefano Rizzi

Urban Analytics of Big Transportation Data for Supporting Smart Cities 24
Carson K. Leung, Peter Braun, Calvin S. H. Hoi, Joglas Souza, and Alfredo Cuzzocrea

Patterns

Frequent Item Mining When Obtaining Support Is Costly 37
Joe Wing-Ho Lin and Raymond Chi-Wing Wong

Mining Sequential Patterns of Historical Purchases for E-commerce Recommendation 57
Raj Bhatta, C. I. Ezeife, and Mahreen Nasir Butt

Discovering and Visualizing Efficient Patterns in Cost/Utility Sequences 73
Philippe Fournier-Viger, Jiaxuan Li, Jerry Chun-Wei Lin, and Tin Truong-Chi

Efficient Row Pattern Matching Using Pattern Hierarchies for Sequence OLAP 89
Yuya Nasu, Hiroyuki Kitagawa, and Kosuke Nakabasami

Statistically Significant Discriminative Patterns Searching 105
Hoang Son Pham, Gwendal Virlet, Dominique Lavenier, and Alexandre Termier

RDF and Streams

Multidimensional Integration of RDF Datasets 119
Jam Jahanzeb Khan Behan, Oscar Romero, and Esteban Zimányi

RDFPartSuite: Bridging Physical and Logical RDF Partitioning 136
Jorge Galicia, Amin Mesmoudi, and Ladjel Bellatreche

Mining Quantitative Temporal Dependencies Between Interval-Based Streams 151
Amine El Ouassouli, Lionel Robinault, and Vasile-Marian Scuturici

Democratization of OLAP DSMS 166
Carlos Garcia-Alvarado, Joy Kent, Li Liu, and Jay Hum

Big Data Systems

Leveraging the Data Lake: Current State and Challenges 179
Corinna Giebler, Christoph Gröger, Eva Hoos, Holger Schwarz, and Bernhard Mitschang

SDWP: A New Data Placement Strategy for Distributed Big Data Warehouses in Hadoop 189
Yassine Ramdane, Nadia Kabachi, Omar Boussaid, and Fadila Bentayeb

Improved Programming-Language Independent MapReduce on Shared-Memory Systems 206
Erik G. Selin and Herna L. Viktor

Evaluating Redundancy and Partitioning of Geospatial Data in Document-Oriented Data Warehouses 221
Marcio Ferro, Rinaldo Lima, and Robson Fidalgo

Graphs and Machine Learning

Scalable Least Square Twin Support Vector Machine Learning 239
Bakshi Rohit Prasad and Sonali Agarwal

Finding Strongly Correlated Trends in Dynamic Attributed Graphs 250
Philippe Fournier-Viger, Chao Cheng, Zhi Cheng, Jerry Chun-Wei Lin, and Nazha Selmaoui-Folcher

Text-Based Event Detection: Deciphering Date Information Using Graph Embeddings 266
Hilal Genc and Burcu Yilmaz

Efficiently Computing Homomorphic Matches of Hybrid Pattern Queries on Large Graphs 279
Xiaoying Wu, Dimitri Theodoratos, Dimitrios Skoutas, and Michael Lan

Databases

From Conceptual to Logical ETL Design Using BPMN and Relational Algebra 299
Judith Awiti, Alejandro Vaisman, and Esteban Zimányi

Accurate Aggregation Query-Result Estimation and Its Efficient Processing on Distributed Key-Value Store 310
Kosuke Yuki, Atsushi Keyaki, Jun Miyazaki, and Masahide Nakamura

Author Index 321