Communications in Computer and Information Science

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
Preface

The Database and Expert Systems Applications (DEXA) workshops are a platform for the exchange of ideas, experiences, and opinions among theoreticians and practitioners – those who are defining requirements for future systems in the areas of database and artificial technologies.

The DEXA workshop papers include papers that are primarily concerned with very specialized topics on applications of database and expert systems technology. We want to thank all workshop organizers for their excellent work.

This was the first time that DEXA workshop papers were published in the *Communications in Computer and Information Science* (CCIS) by Springer.

DEXA 2018 was the 29th annual scientific platform on Database and Expert Systems Applications after Vienna, Berlin, Valencia, Prague, Athens, London, Zurich, Toulouse, Vienna, Florence, Greenwich, Munich, Aix en Provence, Prague, Zaragoza, Copenhagen, Krakow, Regensburg, Turin, Linz, Bilbao, Toulouse, Vienna, Prague, Munich, Valencia, Porto, and Lyon. This year DEXA took place at the University of Regensburg, Germany. Special thanks to Günther Pernul and his local organizing team from the University of Regensburg for hosting DEXA 2018. We would like to express our thanks to all institutions actively supporting this event, namely:

- University of Regensburg
- City of Regensburg
- DEXA Association
- Institute for Application Oriented Knowledge Processing, University of Linz (FAW)
- FAW GmbH

The workshops took place in parallel to the DEXA conference and the program included the following three workshops:

- BDMICS 2018: Third International Workshop on Big Data Management in Cloud Systems
- BIOKDD 2018: 9th International Workshop on Biological Knowledge Discovery from Data
- TIR 2018: 15th International Workshop on Technologies for Information Retrieval

June 2018

A Min Tjoa

Roland Wagner
Organization

General Chairs
Abdelkader Hameurlain  IRIT, Paul Sabatier University Toulouse, France
Günter Pernul  University of Regensburg, Germany
Roland R. Wagner  Johannes Kepler University Linz, Austria

Conference Program Chairs
Hui Ma  Victoria University of Wellington, New Zealand
Sven Hartmann  Clausthal University of Technology, Germany

Workshop Chairs
A Min Tjoa  Technical University of Vienna, Austria
Roland R. Wagner  FAW, University of Linz, Austria
Contents

Big Data Management in Cloud Systems (BDMICS)

Parallel Data Management Systems, Consistency and Privacy

A Survey on Parallel Database Systems from a Storage Perspective: Rows Versus Columns .................................................. 5
Carlos Ordonez and Ladjel Bellatreche

ThespisDIIP: Distributed Integrity Invariant Preservation ............... 21
Carl Camilleri, Joseph G. Vella, and Vitezslav Nezval

Privacy Issues for Cloud Systems .............................................. 38
Christopher Horn and Marina Tropmann-Frick

Cloud Computing and Graph Queries

Script Based Migration Toolkit for Cloud Computing Architecture in Building Scalable Investment Platforms .............................. 46
Rao Casturi and Rajshekhar Sunderraman

Space-Adaptive and Workload-Aware Replication and Partitioning for Distributed RDF Triple Stores ........................................... 65
Ahmed Al-Ghezi and Lena Wiese

Performance Comparison of Three Spark-Based Implementations of Parallel Entity Resolution .................................................. 76
Xiao Chen, Kirity Rapuru, Gabriel Campero Durand, Eike Schallehn, and Gunter Saake

Big Data Analytics: Exploring Graphs with Optimized SQL Queries ........ 88
Sikder Tahsin Al-Amin, Carlos Ordonez, and Ladjel Bellatreche

Biological Knowledge Discovery from Big Data (BIOKDD)

New Modeling Ideas for the Exact Solution of the Closest String Problem ................................................................. 105
Marcello Dalpasso and Giuseppe Lancia

Ensemble Clustering Based Dimensional Reduction ................................................. 115
Loai Abddallah and Malik Yousef
Detecting Low Back Pain from Clinical Narratives Using Machine Learning Approaches ........................................ 126
  Michael Judd, Farhana Zulkernine, Brent Wolfrom, David Barber, and Akshay Rajaram

Classifying Big DNA Methylation Data: A Gene-Oriented Approach ............ 138
  Emanuel Weitschek, Fabio Cumbo, Eleonora Cappelli, Giovanni Felici, and Paola Bertolazzi

Classifying Leukemia and Gout Patients with Neural Networks ............... 150
  Guryash Bahra and Lena Wiese

Incremental Wrapper Based Random Forest Gene Subset Selection for Tumor Discrimment ........................................... 161
  Alia Fatima, Usman Qamar, Saad Rehman, and Aiman Khan Nazir

Protein Identification as a Suitable Application for Fast Data Architecture .... 168
  Roman Zoun, Gabriel Campero Durand, Kay Schallert, Apoorva Patrikar, David Bronkes, Wolfram Fenske, Robert Heyer, Dirk Benndorf, and Gunter Saake

Mining Geometrical Motifs Co-occurrences in the CMS Dataset ................ 179
  Mirto Musci and Marco Ferretti

Suitable Overlapping Set Visualization Techniques and Their Application to Visualize Biclustering Results on Gene Expression Data ............. 191
  Haithem Aouabed, Rodrigo Santamaria, and Mourad Elloumi

Technologies for Information Retrieval (TIR)

Web and Domain Corpora

Can We Quantify Domainhood? Exploring Measures to Assess Domain-Specificity in Web Corpora ........................................ 207
  Marina Santini, Wiktor Strandqvist, Mikael Nyström, Marjan Alirezai, and Arne Jönsson

A Case Study of Closed-Domain Response Suggestion with Limited Training Data .................................................. 218
  Lukas Galke, Gunnar Gerstenkorn, and Ansgar Scherp

What to Read Next? Challenges and Preliminary Results in Selecting Representative Documents ........................................ 230
  Tilman Beck, Falk Böschen, and Ansgar Scherp
NLP Applications

Text-Based Annotation of Scientific Images Using Wikimedia Categories . . . 243
  Frieda Josi, Christian Wartena, and Jean Charbonnier

Detecting Link and Landing Page Misalignment in Marketing Emails . . . 254
  Nedim Lipka, Tak Yeon Lee, and Eunyee Koh

Toward Validation of Textual Information Retrieval Techniques
for Software Weaknesses . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 265
  Jukka Ruohonen and Ville Leppänen

Social Media and Personalization

Investigating the Effect of Attributes on User Trust in Social Media . . . . 278
  Jamal Al Qundus and Adrian Paschke

Analysing Author Self-citations in Computer Science Publications . . . . 289
  Tobias Milz and Christin Seifert

A Semantic-Based Personalized Information Retrieval Approach Using
a Geo-Social User Profile. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 301
  Tahar Rafa and Samir Kechid

Author Index . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 315