DBKDA 2015

The Seventh International Conference on Advances in Databases, Knowledge, and Data Applications

GraphSM 2015

The Second International Workshop on Large-scale Graph Storage and Management

May 24 - 29, 2015

Rome, Italy

DBKDA 2015 Editors

Friedrich Laux, Reutlingen University, Germany

Andreas Schmidt, Karlsruhe University of Applied Sciences | Karlsruhe Institute of Technology, Germany

Maria Del Pilar Angeles, Universidad Nacional Autonoma de Mexico - Del Coyoacan, Mexico

Kiyoshi Nitta, Yahoo Japan Research, Japan

Iztok Savnik, Jozef Stefan Institute and University of Primorska, Slovenia

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571

Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2015) by International Academy, Research, and Industry Association (IARIA)
Please refer to the Copyright Information page.

Printed by Curran Associates, Inc. (2015)

International Academy, Research, and Industry Association (IARIA)
412 Derby Way
Wilmington, DE 19810

Phone: (408) 893-6407
Fax: (408) 527-6351

petre@iaria.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2634
Email: curran@proceedings.com
Web: www.proceedings.com

TABLE OF CONTENTS

DBKDA 1: DATA QUERY, ACCESS, MINING, AND CORRELATION I

An Industry Experience on Dynamic Handling of Multiple Database Types for Data Integration Tool 1
Badrul Affandy Bin Ahmad Latfi, Anbarasan Kodandaraman, Chee Kiam Lee, Tong Khin Thong

Clustering of Words in Texts Using Fuzzy Neighborhood 5
CanLun Zhang, Sadaaki Miyamoto

Improving Near Duplicate Data Detection via DSound Phonetic Matching Algorithm: A Solution to Address Typographical Problems 9
Cihan Varol, Sairam Hari

Query-Based ℓ-Diversity 15
Chittaphone Phonharath, Ryunosuke Takayama, Kenji Hashimoto, Hiroyuki Seki

DBKDA 2: DATABASES AND OTHER DOMAINS

Exposing the Myth: Object-Relational Impedance Mismatch is a Wicked Problem 21
Christopher Ireland, David Bowers

Rule-Based Adaptation of Workflow Patterns for Generic Workflows 27
Marina Tropmann-Frick, Niklas Sasse

JAebXR: a Java API for ebXML Registries for Federated Health Information Systems 33
Antonio Messina, Pietro Storniolo, Alfonso Urso

On the Detection of Nontrivial and Cross Language Plagiarisms 40
Andreas Schmidt, Sören Bühler

DBKDA 3: DATABASES TECHNOLOGIES

A New Spatial Database Management System Using an Hyperbolic Tree 43
Telesphore Tiendrebeogo

Extending PostgreSQL with Column Store Indexes 51
Minoru Nakamura, Tsugichika Tabaru, Yoshifumi Ujibashi, Takushi Hashida, Motoyuki Kawaba, Lilian Harada

A Proposed Architecture to Support the Processing and Analysis of Structured and Unstructured Massive Data Sets in the Brazilian Army 53
Marcio Victorino, Marçal Hokama, Danilo Silva

DBKDA 4: DATA MANAGEMENT

Analysis of String Comparison Methods During De-Duplication Process 57
Maria del Pilar Angeles, Francisco García-Ugalde, Ricardo Valencia, Arturo Nava

Comparison of Methods Hamming Distance, Jaro, and Monge-Elkan 63
Maria del Pilar Angeles, Adrian Espino-Gamez

An Efficient Approach to Triple Search and Join of HDT Processing Using GPU 70
YoonKyung Kim, YoonJoon Lee, JaeHwan Lee

Towards a New Incremental Implementation Approach for Distributed Databases 75
Marwa Bouraoui, Fadoua Hassen, Grissa touzi Amel

DBKDA 5: KNOWLEDGE AND DECISION BASES

OBDBSearch: A Keyword-based Semantic Search for Databases 82
Simone Miraglia, Clarissa Molfino, Costanza Romano

Automatic Knowledge Transfer-based Architecture towards Self-Service Business Intelligence 90
Safwan Sulaiman, Tariq Mahmoud, Jorge Marx Gómez, Joachim Kurzhöfer

IES's UIEs for Generating LACs for Arguing that a CI Satisfies SPL 97
Sigram Schindler

Applying Fuzzy Weights to Triple Inner Dependence AHP 104
Shin-ichi Ohnishi, Takahiro Yamanoi

DBKDA 6: DATA QUERY, ACCESS, MINING, AND CORRELATION II

Near Real-time Synchronization Approach for Heterogeneous Distributed Databases 107
Hassen Fadoua, Grissa Touzi Amel
Skyline Objectset: Efficient Selection of Non-dominate Sets from Database 114
Anisuzzaman Siddique, Asif Zaman, Yasuhiko Morimoto
On the Number of Rules and Conditions in Mining Incomoplete Data with Lost Values and Attribute-concept Values..... 121
Jerzy W. Grzymala-Busse, Patrick G. Clark
Patent Threat Analysis Search Engine..... 127
Yung Chang Chi, Hei Chia Wang

DBKDA 7: DATABASES CONTENT PROCESSING I

Dataconda -- A Software Framework for Mining Relational Databases 132
Michele Samorani
Efficient Media Digital Library of Summarized Video based on Scalable Video Coding for H.264 (MDLSS): Design and Implementation..... 134
Hesham Farouk, Mayada Khairy, Kamal El-Dahshan, Amr Abozeid, Alaa Hamdy, Amr Elsayed
Behind the Skyline 141
Markus Endres, Timotheus Preisinger
ORPE - A Data Semantics Driven Concurrency Control Mechanism 147
Tim Lessner, Fritz Laux, Thomas M Connollly

DBKDA 8: DATABASES CONTENT PROCESSING II

Towards Fuzzy Querying of NoSQL Document-oriented Databases 153
Abir Belhadj Kacem, Amel Grissa
Combining Distributed Computing and Massively Parallel Devices to Accelerate Stream Data Processing..... 159
David Bednarek, Martin Krulis, Petr Maly, Jakub Yaghob, Filip Zavoral, Jaroslav Pokorny
Accelerating Data Mining on Incomplete Datasets by Bitmaps-based Missing Value Imputation..... 167
Sameh Shohdy, Yu Su, Gagan Agrawal

GRAPHSM 1: GRAPH STORAGE AND MANAGEMENT

Capturing the Structure of Internet of Things Systems with Graph Databases..... 176
Gilles Privat, Dana Popovici
Towards Implementing Semantic Literature-Based Discovery with a Graph Database 180
Dimitar Hristovski, Andrej Kastrin, Dejan Dinevski, Thomas C. Rindflesch
Author Index