



# **ALLDATA 2017**

The Third International Conference on Big Data, Small Data, Linked Data and Open  
Data

## **KESA 2017**

The International Workshop on Knowledge Extraction and Semantic Annotation

April 23 - 27, 2017

Venice, Italy

## **ALLDATA 2017 Editors**

Venkat N. Gudivada, East Carolina University, USA

Sandjai Bhulai, Vrije Universiteit Amsterdam, the Netherlands

Maria Pia di Buono, University of Zagreb, Croatia

**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© (2017) by International Academy, Research, and Industry Association (IARIA)  
Please refer to the Copyright Information page.

Printed by Curran Associates, Inc. (2017)

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

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

[petre@iaria.org](mailto: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: 845-758-2633  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

# TABLE OF CONTENTS

## **ALLDATA 1: BIG DATA**

<b>Flexible Management of Data Nodes for Hadoop Distributed File System</b> .....	1
<i>W. Ryu</i>	
<b>Curves Similarity Based on Higher Order Derivatives</b> .....	3
<i>F. Nicol, S. Puechmore</i>	
<b>Automated Generation of SQL Queries that Feature Specified SQL Constructs</b> .....	9
<i>V. Gudivada, K. Arbabifard, D. Rao</i>	

## **ALLDATA 2: LINKED DATA**

<b>Data Visualization of A Cyber-Physical Systems Development Toolchain: An Integration Case Study</b> .....	14
<i>D. Gurdur, J. El-Khoury, T. Seceleanu, M. Johansson, S. Hansen</i>	
<b>Processing Information about Junior Specialists for Small IT-projects Teams Using Linked Data</b> .....	20
<i>N. Yussupova, O. Smetanina, D. Rizyanov, A. Galyamov</i>	

## **ALLDATA 3: OPEN DATA**

<b>Reproducible Evaluation of Semantic Storage Options</b> .....	26
<i>J. Rybicki, V. Vieth</i>	
<b>TLEX: A Temporal Analysis Tool for Time Series Data</b> .....	30
<i>M. Zamil, B. Huda</i>	
<b>Empirical Evaluation of Open Government Data Visualisations</b> .....	35
<i>E. Ormig, J. Faichney, B. Stantic</i>	

## **ALLDATA 4: CHALLENGES IN PROCESSING BIG DATA AND APPLICATIONS I**

<b>Analyzing Browsing and Purchasing Across Multiple Websites Based on Latent Dirichlet Allocation</b> .....	40
<i>N. Schroder, A. Falke, H. Hruschka, T. Reutterer</i>	

## **ALLDATA 5: CHALLENGES IN PROCESSING BIG DATA AND APPLICATIONS II**

<b>A Vision on Prescriptive Analytics</b> .....	45
<i>M. Sappelli, S. Smit, M. Boer, F. Bomhof</i>	
<b>Statistical Analysis of Aircraft Trajectories: a Functional Data Analysis Approach</b> .....	51
<i>F. Nicol</i>	

## **KESA 1: DATA MINING AND KNOWLEDGE EXTRACTION**

<b>Analyzing Characteristics of Picture Books based on an Infant’s Developmental Reactions in Reviews on Picture Books</b> .....	57
<i>M. Baba, H. Uehara, M. Kasamatsu, T. Utsuro, C. Zhao</i>	
<b>A Rule-based Named Entity Extraction Method and Syntactico-Semantic Annotation for Arabic Language</b> .....	63
<i>L. Chahira, Z. Anis, Z. Mounir</i>	

## **KESA 2: ANALYSIS OF ONTOLOGY MODELS FOR NATURAL LANGUAGE TEXTS**

<b>iTweet about #Privacy - Mapping Privacy Frames in Twitter Conversation</b> .....	70
<i>F. Fornaciari</i>	

<b>Industry Experience: Chinese Names Duplicate Records Detection</b> .....	74
<i>T. Khin, B. Ahmad, W. Xiaomei, K. Arichandran</i>	
<b>Topic Models to Contextualize and Enhance Text-Based Discourses Using Ontologies</b> .....	78
<i>D. Gkoumas, R. Vas</i>	
<b>Author Index</b>	