

**PROCEEDINGS OF  
THE 2014 INTERNATIONAL CONFERENCE ON  
ADVANCES IN BIG DATA ANALYTICS**

# **ABDA 2014**

## **Editors**

**Hamid R. Arabnia  
Quoc-Nam Tran**

## **Associate Editors**

**George Jandieri, James J. (Jong Hyuk) Park  
Ashu M. G. Solo  
Fernando G. Tinetti**



***WORLD COMP'14***

July 21-24, 2014

Las Vegas Nevada, USA

[www.world-academy-of-science.org](http://www.world-academy-of-science.org)

©CSREA Press

This volume contains papers presented at The 2014 International Conference on Advances in Big Data Analytics (ABDA'14). Their inclusion in this publication does not necessarily constitute endorsements by editors or by the publisher.

### **Copyright and Reprint Permission**

Copying without a fee is permitted provided that the copies are not made or distributed for direct commercial advantage, and credit to source is given. Abstracting is permitted with credit to the source. Please contact the publisher for other copying, reprint, or republication permission.

Copyright © 2014 CSREA Press  
ISBN: 1-60132-264-X  
Printed in the United States of America

CSREA Press  
U. S. A.

# Foreword

It gives us great pleasure to introduce this collection of papers to be presented at the 2014 International Conference on Advances in Big Data Analytics (ABDA'14), July 21-24, 2014, at Monte Carlo Resort, Las Vegas, USA.

An important mission of the World Congress in Computer Science, Computer Engineering, and Applied Computing (a federated congress to which this conference is affiliated with) includes *"Providing a unique platform for a diverse community of constituents composed of scholars, researchers, developers, educators, and practitioners. The Congress makes concerted effort to reach out to participants affiliated with diverse entities (such as: universities, institutions, corporations, government agencies, and research centers/labs) from all over the world. The congress also attempts to connect participants from institutions that have **teaching** as their main mission with those who are affiliated with institutions that have **research** as their main mission. The congress uses a quota system to achieve its institution and geography diversity objectives."* By any definition of diversity, this congress is among the most diverse scientific meeting in USA. We are proud to report that this federated congress has authors and participants from 78 different nations representing variety of personal and scientific experiences that arise from differences in culture and values. As can be seen (see below), the program committee of this conference as well as the program committee of all other tracks of the federated congress are as diverse as its authors and participants.

The program committee would like to thank all those who submitted papers for consideration. About 58% of the submissions were from outside the United States. Each submitted paper was peer-reviewed by two experts in the field for originality, significance, clarity, impact, and soundness. In cases of contradictory recommendations, a member of the conference program committee was charged to make the final decision; often, this involved seeking help from additional referees. In addition, papers whose authors included a member of the conference program committee were evaluated using the double-blinded review process. One exception to the above evaluation process was for papers that were submitted directly to chairs/organizers of pre-approved sessions/workshops; in these cases, the chairs/organizers were responsible for the evaluation of such submissions. The overall paper acceptance rate for regular papers was 27; 15% of the remaining papers were accepted as poster papers (at the time of this writing, we had not yet received the acceptance rate for a few individual tracks.)

We are very grateful to the many colleagues who offered their services in organizing the conference. In particular, we would like to thank the members of the Program Committee of ABDA'14, members of the congress Steering Committee, and members of the committees of federated congress tracks that have topics within the scope of ABDA. Many individuals listed below, will be requested after the conference to provide their expertise and services for selecting papers for publication (extended versions) in journal special issues as well as for publication in a set of research books (to be prepared for publishers including: Springer, Elsevier, BMC, and others).

- Prof. Afrand Agah; Assistant Chair & Graduate Coordinator, Department of Computer Science, West Chester University of Pennsylvania, West Chester, Pennsylvania, USA
- Dr. Selim Aissi (Congress Steering Committee); Vice President, Global Information Security, Visa Inc., USA (formerly: Chief Strategist - Security, Intel Corporation, USA)
- Dr. Eng. Hamid Ali Abed Al-Asadi (ICOMP); Head, Computer Science Department, Faculty of Education for Pure Science, Basra University, Basra, Iraq
- Prof. Rafeeq AbdulRahman Al-Hashemi; Dean of Information Technology, Al-Hussein Bin Talal University, Jordan
- Prof. Nizar Al-Holou (Congress Steering Committee); Professor and Chair, Electrical and Computer Engineering Department; Vice Chair, IEEE/SEM-Computer Chapter; University of Detroit Mercy, Detroit, Michigan, USA
- Prof. Hamid R. Arabnia (Congress Steering Committee); Professor of Computer Science; The University of Georgia, Graduate Studies Research Center, Athens, Georgia, USA; Editor-in-Chief, Journal of Supercomputing (Springer); Editor-in-Chief, Emerging Trends in CS and Applied Computing (Elsevier); Editor-in-Chief, Transactions of Computational Science & Computational Intelligence (Springer); Elected Fellow, Int'l Soc. of Intelligent Biological Medicine (ISIBM); USA

- Prof. Ezendu Ariwa; Faculty of Computer Science and Technology, University of Bedfordshire, UK; Chair, IEEE Consumer Electronics & Broadcast Technology Chapter, UKRI; Chair, IEEE Technology Management Council Chapter, UKRI, UK
- Mehdi Bahrami; IEEE Senior Member, Senior Software Analyst and Software Manager, Lian Processor Co.; Cloud Lab, Electrical Engineering and Computer Science Department (EECS), University of California, Merced, USA; BISC member, University of California, Berkeley, California, USA
- Prof. Michael Panayiotis Bekakos (Congress Steering Committee); Professor of Computer Systems; Director, Laboratory of Digital Systems, Department of Electrical and Computer Engineering, Democritus University of Thrace, Greece; Head, Parallel Algorithms and architectures Research Group (PAaRG); Editor-in-Chief, The Journal of Neural, Parallel & Scientific Computations, USA
- Dr. Sidahmed Benabderrahmane; INRIA (French National Computer Science Institute), France
- Prof. Juan-Vicente Capella-Hernandez; Executive and Quality Manager, Wireless Sensor Networks Valencia, Inc.; Member, Editorial Board: IEEE RITA Journal; International Journal On Advances in Networks and Services; International Journal of Computer Science & Information Technology Applications; Universitat Politècnica de Valencia, Valencia, Spain
- Prof. Juan Jose Martinez Castillo; Director of The Acantelys Research Group and Coordinator of the Computer Engineering Department, Universidad Gran Mariscal de Ayacucho, Venezuela
- Dr. Daniel Bo-Wei Chen; Department of Electrical Engineering, Princeton University, New Jersey, USA
- Dr. Xin Chen (ICOMP); Research Corporation of the University of Hawaii / Institute for Astronomy, Honolulu, Hawaii, USA
- Prof. Kevin Daimi (Congress Steering Committee); Director, Computer Science and Software Engineering Programs, Department of Mathematics, Computer Science and Software Engineering, University of Detroit Mercy, Detroit, Michigan, USA
- Somdip Dey (ICOMP); School of Computer Science, The University of Manchester, Manchester, UK
- Dr. Lamia Atma Djoudi; Synchrone Technologies, France
- Prof. Mary Mehrnoosh Eshaghian-Wilner (Congress Steering Committee); Professor of Engineering Practice, University of Southern California, California, USA; Adjunct Professor, Electrical Engineering, University of California Los Angeles, Los Angeles (UCLA), California, USA
- Prof. Mohammad Shahadat Hossain; Department of Computer Science and Engineering, University of Chittagong, Chittagong, Bangladesh; Visiting Professor, Trisakti University, Indonesia; Visiting Academic Staff, The University of Manchester, UK
- Dr. Guofeng Hou (ICOMP); Bell Laboratories and Microsoft Corporation, USA
- Prof. George Jandieri (Congress Steering Committee & Vice Chair of CSC); Georgian Technical University, Tbilisi, Georgia; Chief Scientist, The Institute of Cybernetics, Georgian Academy of Science, Georgia; Editorial Board Member: International Journal of Microwaves and Optical Technology, The Open Atmospheric Science Journal, American Journal of Remote Sensing
- Prof. Young-Sik Jeong (Congress Steering Committee); Department of Multimedia Engineering, Dongguk University, Seoul, South Korea; Editor-in-Chief, Journal of Information Processing Systems (JIPS)
- Prof. Byung-Gyu Kim (Congress Steering Committee); Multimedia Processing Communications Lab.(MPCL), Department of Computer Science and Engineering, College of Engineering, SunMoon University, South Korea
- Prof. Dattatraya V. Kodavade (Congress Steering Committee); Head of Computer Science and Engineering, D.K.T.E Society's Textile & Engineering Institute, Ichalkaranji, Maharashtra State, India
- Gurudatt Anil Kulkarni (ICOMP); I/C Head of Department and Lecturer, Electronics & Telecommunication Department, Marathwada Mitra Mandal's Polytechnic, Pune, India
- Prof. George Markowsky (Congress Steering Committee); Professor and Associate Director, School of Computing and Information Science; Chair International Advisory Board of IEEE IDAACS; Director 2013 Northeast Collegiate Cyber Defense Competition; Chair Bangor Foreign Policy Forum; President Phi Beta Kappa Delta Chapter of Maine; Cooperating Prof. Mathematics & Statistics Department UMaine; Cooperating Prof. School of Policy & Int'l Affairs UMaine; University of Maine, Orono, Maine, USA
- Dr. Andrew Marsh (Congress Steering Committee); CEO, HoIP Telecom Ltd (Healthcare over Internet Protocol), UK; Secretary General of World Academy of BioMedical Sciences and Technologies (WABT) a UNESCO NGO, The United Nations
- Dr. Ben C.K. Ngan; The Pennsylvania State University, Great Valley School of Graduate Professional Studies, Division of Engineering and Information Science, Malvern, Pennsylvania, USA; Adjunct Professor, Department of Applied Information Technology, George Mason University, Fairfax, Virginia, USA
- Prof. G. N. Pandey (Congress Steering Committee); Vice-Chancellor, Arunachal University of Studies, Arunachal Pradesh, India; Adjunct Professor, Indian Institute of Information Technology, Allahabad, India
- Prof. James J. (Jong Hyuk) Park (Congress Steering Committee); Department of Computer Science and Engineering (DCSE), SeoulTech, Korea; President, FTRA, Editor-in-Chief, HCIS Springer, JoC, IJITCC; Head of DCSE, SeoulTech, Korea

- *Dr. Xinyu Que; IBM T. J. Watson Research Center, Yorktown Heights, New York, USA*
- *Ashu M. G. Solo (Congress Steering Committee & Publicity Chair); Fellow of British Computer Society, Principal/R&D Engineer, Maverick Technologies America Inc.*
- *Prof. Fernando G. Tinetti (Congress Steering Committee); School of Computer Science, Universidad Nacional de La Plata, La Plata, Argentina; Co-editor, Journal of Computer Science and Technology (JCS&T)*
- *Prof. Patrick S. P. Wang (Congress Steering Committee); Fellow: IAPR, ISIBM, WASE; Professor of Computer and Information Science, Northeastern University, Boston, Massachusetts, USA and Zijiang Visiting Chair, ECNU, Shanghai, NTUST, Taipei; iCORE Visiting Professor, University of Calgary, Canada; Otto-von-Guericke Distinguished Guest Professor, University Magdeburg, Germany*
- *Prof. Shih-Jeng Wang (Congress Steering Committee); Department of Information Management, Central Police University, Taiwan; Program Chair, Security & Forensics, Taiwan; R.O.C.; Director, Information Crypto and Construction Lab (ICCL) & ICCL-FROG*
- *Dr. Wei Wei (ICOMP); School of Computer Science and Engineering, Xi'an University of Technology, Xi'an, P. R. China; Member of China Computer Federation (CCF), P. R. China*
- *Dr. Alexander Woehrer; Program Manager ICT, Vienna Science and Technology Fund, Vienna, Austria*
- *Prof. Mary Q. Yang (Congress Steering Committee and Vice Chair of ABDA and BIOCOMP); Director, Mid-South Bioinformatics Center and Joint Bioinformatics Ph.D. Program, Medical Sciences and George W. Donaghey College of Engineering and Information Technology, University of Arkansas, USA*

We would like to extend our appreciation to the members of the program committees of individual sessions, tracks, and workshops; their names do not appear in this document; they are listed on the web sites of individual tracks.

As Sponsors-at-large, partners, and/or organizers each of the followings (separated by semicolons) provided help for at least one track of the World Congress: Computer Science Research, Education, and Applications Press (CSREA); Taylor & Francis, UK (<http://www.taylorandfrancis.com/>); US Chapter of World Academy of Science (<http://www.world-academy-of-science.org/>); American Council on Science & Education & Federated Research Council (<http://www.americancse.org/>); HoIP, Health Without Boundaries, Healthcare over Internet Protocol, UK (<http://www.hoip.eu>); HoIP Telecom, UK (<http://www.hoip-telecom.co.uk>); and WABT, Human Health Medicine, UNESCO NGOs, Paris, France (<http://www.thewabt.com/>). In addition, a number of university faculty members and their staff (names appear on the cover of the set of proceedings), several publishers of computer science and computer engineering books and journals, chapters and/or task forces of computer science associations/organizations from 5 countries, and developers of high-performance machines and systems provided significant help in organizing the conference as well as providing some resources. We are grateful to them all.

We express our gratitude to keynote, invited, and individual conference/tracks and tutorial speakers - the list of speakers appears on the conference web site. We would also like to thank the followings: UCMSS (Universal Conference Management Systems & Support, California, USA) for managing all aspects of the conference; Dr. Tim Field of APC for managing the printing of the proceedings; and the staff of Monte Carlo Resort (Convention department) in Las Vegas for the professional service they provided. Last but not least, we would like to thank the Co-Editors and Associate Co-Editors of ABDA'14: Prof. Hamid R. Arabnia, Prof. George Jandieri, Prof. James J. (Jong Hyuk) Park, Ashu M. G. Solo, Prof. Fernando G. Tinetti, and Prof. Mary Q. Yang.

We present the proceedings of ABDA'14.

***Steering Committee, 2014***

***<http://www.world-academy-of-science.org/>***



# Contents

## SESSION: APPLICATIONS OF BIG DATA + SOFTWARE TOOLS

<b>Application-Agnostic Streaming Bayesian Inference via Apache Storm</b>	<b>3</b>
<i>Todd Wasson, Ana Paula de Oliveira Sales</i>	
<b>Multicore Construction of k-d Trees for High Dimensional Point Data</b>	<b>10</b>
<i>Victor Lu, John C. Hart</i>	
<b>Design and Implementation of a Cloud-Based Big Data Programming Service</b>	<b>16</b>
<i>Yen-Yu Lin, Shu-Ming Chang, Su-Shien Ho, Ying-Ti Liao, Hsing-Chang Chou, Shih-Chang Chen, Xuan-Yi Lin, Jiazheng Zhou, Yeh-Ching Chung</i>	
<b>ARIANA: Adaptive Robust and Integrative Analysis for finding Novel Associations</b>	<b>22</b>
<i>Vida Abedi, Mohammed Yeasin, Ramin Zand</i>	
<b>A Software Toolkit for Stock Data Analysis Using Social Network Analysis Approach</b>	<b>29</b>
<i>Junyan Zhang, Donglei Du, Weichang Du</i>	
<b>A Visual Decision-Guided Analytics Tool for Finding the Viable Shortest Path over Geospatiotemporal Data</b>	<b>36</b>
<i>Chun-Kit Ngan</i>	
<b>Using Influence for Navigation in Online Social Networks</b>	<b>44</b>
<i>Bastien Lebayle, Mehran Asadi, Afrand Agah</i>	
<b>Sentiment Analysis for Mobile SNS Data</b>	<b>50</b>
<i>Il-Kyu Ha, Bong-Hyun Back, SeonHwan Kim, Byoungchul Ahn</i>	
<b>E-commerce Evolution based on the Perspective of Network Effects:A Multi-case Study in China</b>	<b>56</b>
<i>Peng Shao, Ping Hu, Jie Qi</i>	
<b>Estimating the Visit and Content Association of Websites with their Weblogs</b>	<b>63</b>
<i>Hyun-Ho Lee, Jong-Min Lee, Nam-Hun Park, Won-Suk Lee</i>	

## SESSION: ALGORITHMS FOR BIG DATA

<b>Noise Benefits in Convolutional Neural Networks</b>	<b>73</b>
<i>Kartik Audhkhasi, Osonde Osoba, Bart Kosko</i>	
<b>On Walsh-Fourier Bispectral Analysis</b>	<b>81</b>
<i>Mahmoud Gabr, Mahmoud El-Hashash, Gehan Sabbah</i>	

<b>Reasearch of Decision Tree on YARN Using MapReduce and Spark</b>	<b>88</b>
<i>Hua Wang, Bin Wu, Shuai Yang, Bai Wang, Yang Liu</i>	

<b>Kernel Approach to Incomplete Data Analysis (KAIDA)</b>	<b>95</b>
<i>S.Y. Kung, Pei-Yuan Wu</i>	

<b>On the usage of Sorting Networks to Big Data</b>	<b>102</b>
<i>Blanca Lopez, Nareli Cruz-Cortes</i>	

<b>Client Based Power Iteration Clustering Algorithm to Reduce Dimensionality in Big Data</b>	<b>108</b>
<i>Jayalatchumy Dhanapal, Thambidurai Perumal</i>	

## **SESSION: BIG DATA MANAGEMENT AND FRAMEWORKS**

<b>Real-Time Data/Information Storage and Retrieval in Autonomous AI Systems</b>	<b>115</b>
<i>James Crowder, John Carbone</i>	

<b>Transparency and Efficiency in Grid Computing for Big Data</b>	<b>121</b>
<i>Paul L. Bergstein</i>	

<b>Integration of Big Data Components for NoSQL Problems</b>	<b>128</b>
<i>Greg Speegle, Erich Baker</i>	

<b>Comparison and Review of Memory Allocation and File Access Techniques and Techniques preferred for Distributed Systems</b>	<b>135</b>
<i>Muazzam A. Khan Khattak, Nauman Nisar</i>	

## **SESSION: BIG DATA SEARCH AND MINING METHODS**

<b>Decomposition of Inverted Lists and Word Labeling: A New Index Structure for Text Search</b>	<b>143</b>
<i>Yangjun Chen, Weixin Shen</i>	

<b>MPGM: A Mixed Parallel Big Graph Mining Tool</b>	<b>151</b>
<i>Pengjiang Ma, Bin Wu, Liu Yang, Hongxu Wang</i>	

<b>Spatial Association Mining of Focal Events in Cloud Computing</b>	<b>157</b>
<i>Sang Jun Park, Jin Soung Yoo</i>	

<b>Spectrum Recognition in Large-Scale Cognitive Radio Networks With Spectral Data Mining</b>	<b>164</b>
<i>Adam Anderson, Charles Witherspoon, Behnaz Papari</i>	



## **SESSION: SECURITY AND PRIVACY IN THE ERA OF BIG DATA AND RISK ANALYSIS**

**Some Risk Measures and Their Applications in Financial Data Analysis** 173

*Qi Tang, Linghua Zhang*

**The Use of Fully Homomorphic Encryption in Data Mining with Privacy Preserving** 180

*Laecio A. Costa, Ruy J. B. de Queiroz*

**Secure Outsourcing Multiparty Computation on Lattice-based Encrypted Cloud Data under Multiple Keys** 187

*Yi Sun, Qiaoyan Wen, Hua Zhang, Zhengping Jin*

## **SESSION: LATE BREAKING PAPERS AND POSITION PAPERS: BIG DATA ANALYTICS AND RELATED ISSUES**

**Querying Clusters of Biological Sequences** 197

*Vicente Molieri, Lina Karam, Zoe Lacroix*

**MapReduce-Accelerated Framework for Identifying Minimum-Sized Influential Vertices on Large-Scale Weighted Graphs** 203

*Ying Xie, Jing He, Vijay Raghavan*

**Lessons Learned: Building a Big Data Research and Education Infrastructure** 209

*George Hsieh, Raymond Sye, Shontae Vincent, Wallace Hendricks*

**Big Data System Design for Digital Library Middleware** 216

*Behrooz Seyed-Abbassi, Chris Carey*

**Finding Critical Samples for Mining Big Data** 222

*Andrew Sung, Bernardete Ribeiro, Qingzhong Liu, Divya Suryakumar*

**Semantic Indexing of Big Data Using a Hierarchical, Multidimensional Scheme** 226

*Ronald Wroblewski*

**Social Event Radar (SERv2.0): Efficient Dynamic Adjustment Mechanism of Distributed Web Crawlers in Social Networks** 233

*Tsun Ku, Cheng-Hung Tsai, Ping-Yen Yang, Ming-Jen Chen*

**Finding Local and Periodic Association Rules from Fuzzy Temporal Data** 240

*Fokrul Mazarbhuiya, Mohamed Shenify, Mohammed Husamuddin*

**Exploring Big-data Analysis using Integrative Systems Biology Approaches for Kidney Renal Clear Cell Carcinoma Studies 248**

*William Yang, Kenji Yoshigoe, Xiang Qin, Jun Liu, A Keith Dunker, Youping Deng, Weida Tong, Andrzej Niemieko, Jack Yang, Yunlong Liu, Zhongxue Chen, Liangjiang Wang, Dong Xu, Hamid Arabnia, Mary Yang*

**Exploring and Analyzing Big Data Processing Technology Based on the EMU 257**

*Qiming Niu, Feng Liu, Jie Liu, Chun Zhang*

**Big Data Architecture for Large-scale Scientific Computing 260**

*Benoit Lange, Toan Nguyen*

**Graph Database for Agent Based Emergency Response Model 264**

*Safiye Sencer, Kutlu Eren*

**Blur Detection For Video Streams In The Compressed Domain 274**

*Zhenyu Wu, Daiying Zhou, Hong Hu*