PROCEEDINGS OF
THE 2015 INTERNATIONAL CONFERENCE ON
ADVANCES IN BIG DATA ANALYTICS

ABDA 2015

Editors

Hamid R. Arabnia
Mary Q. Yang

Associate Editors

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

WORLDCOMP'15
July 27-30, 2015
Las Vegas Nevada, USA
www.world-academy-of-science.org

©CSREA Press

This volume contains papers presented at The 2015 International Conference on Advances in Big Data Analytics (ABDA'15). 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 © 2015 CSREA Press
ISBN: 1-60132-411-1
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 2015 International Conference on Advances in Big Data Analytics (ABDA'15), July 27-30, 2015, 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 76 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 57% 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 29%; 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'15, 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 journals, and others).

- *Prof. Afrand Agah; Assistant Chair and Graduate Coordinator, Department of Computer Science, West Chester University of Pennsylvania, West Chester, Pennsylvania, USA*
- *Assoc. Prof. Dr. Mehmet Fatih Akay, Department of Computer Engineering, Cukurova University, Turkey*
- *Prof. Abbas M. Al-Bakry (Congress Steering Committee); University President, University of IT and Communications, Baghdad, Iraq*
- *Prof. Nizar Al-Holou (Congress Steering Committee); Professor and Chair, ECE Department; Vice Chair, IEEE/SEM-Computer Chapter; University of Detroit Mercy, Detroit, Michigan, USA*
- *Dr. Hamid Ali Abed Alasadi (Congress Steering Committee); Head, Department of Computer Science, Basra University, Iraq; Member of Optical Society of America (OSA), USA; Member of The International Society for Optical Engineering (SPIE), Bellingham, Washington, USA*
- *Prof. Hamid R. Arabnia (Congress Steering Committee & Coordinator & ABDA'15 Co-Editor); Professor of Computer Science; The University of Georgia, USA; Editor-in-Chief, Journal of Supercomputing (Springer); Editor-in-Chief, Emerging Trends in Computer Science and Applied Computing (Elsevier); Editor-in-Chief, Transactions of Computational Science & Computational Intelligence (Springer); Elected Fellow, Int'l Society of Intelligent Biological Medicine (ISIBM); USA*

- *Dr. Mellal Mohamed Arezki (ICAI'15); Associate Professor, Faculty of Engineering Sciences (FSI), M'Hamed Bougara University, Boumerdes, Algeria*
- *Prof. Mehran Asadi; Interim Chair, Department of Business and Entrepreneurial Studies, The Lincoln University, Pennsylvania, USA*
- *Prof. P. Balasubramanian (PDPTA'15); School of Computer Engineering, Nanyang Technological University, Singapore*
- *Prof. Hans-Peter Bischof (PDPTA'15); Chair/Director, Computer Science MS Program and Graduate Program Coordinator, Department of Computer Science, Rochester Institute of Technology, New York, USA*
- *Charles D. Carson; Alumnus Southern Polytechnic State University, Marietta, Georgia, USA*
- *Prof. Juan Jose Martinez Castillo; Director, The Acantelys Alan Turing Nikola Tesla Research Group and GIPEB, Universidad Nacional Abierta, Venezuela*
- *Dr. Bo-Wei Chen; Department of Electrical Engineering, Princeton University, Princeton, New Jersey, USA; and Interim Chair, IEEE Signla Processing Chapter, Harbin Section*
- *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*
- *Dr. Lamia Atma Djoudi (Chair, Doctoral Colloquium); 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. George A. Gravvanis (Congress Steering Committee); Director, Physics Laboratory & Head of Advanced Scientific Computing, Applied Math & Applications Research Group; Professor of Applied Mathematics and Numerical Computing and Department of ECE, School of Engineering, Democritus University of Thrace, Xanthi, Greece; former President of the Technical Commission on Data Processing, Social Security for the Migrant Workers, European Commission, Hellenic Presidency, Greece*
- *Prof. Mohammad Shahadat Hossain (PhD, UMIST, Manchester), MBCS; Department of Computer Science and Engineering, University of Chittagong, Bangladesh; Visiting Academic Staff, The University of Manchester, UK*
- *Prof. George Jandieri (Congress Steering Committee); 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. Byung-Gyu Kim (Congress Steering Committee); Multimedia Processing Communications Lab.(MPCL), Department of Computer Science and Engineering, College of Engineering, SunMoon University, South Korea*
- *Dr. Bo Liu; Spatio-temporal Data Analysis Research Department, NEC Labs China, Beijing, P. R. China (formerly at Argonne National Laboratory, USA)*
- *Dr. Yan Luo; National Institutes of Health, Bethesda, Maryland, USA*
- *Prof. George Markowsky (Congress Steering Committee); Professor & Associate Director, School of Computing and Information Science; Chair International Advisory Board of IEEE IDAACS; Director 2013 Northeast Collegiate Cyber Defense Competition; 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*
- *Prof. Francesc D. Munoz-Escoi (PDPTA'15); E.T.S. Informatica (ETSInf), Departamento de Sistemas Informaticos y Computacion (DSIC), Instituto Universitario Mixto Tecnologico de Informatica (ITI), Universitat Politecnica de Valencia (Polytechnical University of Valencia), Valencia, Spain*
- *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, EiC, HCIS Springer, JoC, IJITCC; Head of DCSE, SeoulTech, Korea*
- *Dr. Alvaro Rubio-Largo (BIOCOMP'15 + ICAI'15); University of Extremadura, Caceres, Spain*
- *Prof. Khemaissia Seddik (ICAI'15 + BIOCOMP'15 + BIOENG'15); University of Tebessa, Algerie, Algeria*
- *Dr. Vijay Bhaskar Semwal (ICAI'15 + IKE'15); Department of Robotics & Artificial Intelligence, Indian Institute of Information Technology (IIIT), Allahabad, India*
- *Dr. Akash Singh (Congress Steering Committee); IBM Corporation, Sacramento, California, USA; Chartered Scientist, Science Council, UK; Fellow, British Computer Society; Member, Senior IEEE, AACR, AAAS, and AAAI; IBM Corporation, USA*

- *Ashu M. G. Solo, (Publicity Chair), Fellow of British Computer Society, Principal/R&D Engineer, Maverick Technologies America Inc.*
- *Asst. Prof. K. Suresh (PDPTA'15); Department of Information Technology, Annamacharya Institute of Technology and Sciences, (Autonomous), Rajampet, A.P., India*
- *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 & Technology (JCS&T)*
- *Mohd Helmy Abd Wahab; Senior Lecturer and Academic Advisor, Department of CE, Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia (UTHM), Johor, Malaysia*
- *Prof. Shiuh-Jeng Wang (Congress Steering Committee); Department of Information Management, Central Police University, Taiwan; Program Chair, Security & Forensics, Taiwan; Director, Information Crypto and Construction Lab (ICCL) & ICCL-FROG*
- *Prof. Dr. Bernd E. Wolfinger; Telecommunications and Computer Networks Division, Computer Science Department, University of Hamburg, Hamburg (Stellingen), Germany*
- *Prof. Mary Q. Yang (Congress Steering Committee & ABDA'15 Vice Chair & Co-Editor); 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*
- *Prof. Jane You (Congress Steering Committee); Associate Head, Department of Computing, The Hong Kong Polytechnic University, Kowloon, Hong Kong*
- *Dr. Nezer Jacob Zaidenberg (ABDA'15 Session Organizer); Shenkar College of Engineering and Design, Department of Software Engineering, Ramat Gan, Israel; University of Jyvakyla, Department of Mathematical Information Technology, Finland; and Adjunct Prof., Tel Aviv Jaffa Academic College, Israel*
- *Dr. Fang Zheng (PDPTA'15 + GCA'15); IBM T.J. Watson Research Center, Yorktown Heights, NY, 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); 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 4 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 coordinating and 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'15: Prof. Hamid R. Arabnia, Prof. George Jandieri, Prof. James J. (Jong Hyuk) Park, Ashu M. G. Solo, Prof. Fernando G. Tinetti, and Dr. Mary Q. Yang.

We present the proceedings of ABDA'15.

Steering Committee, 2015
http://www.world-academy-of-science.org/

Contents

SESSION: HIGH PERFORMANCE COMPUTING, DISTRIBUTED AND PARALLEL PROCESSING, DATABASES

Benchmarking and Performance Studies of Mapreduce, Hadoop Framework on Blue Waters Supercomputer **3**
Manisha Gajbe, Kalyana Chadalavada, Gregory Bauer, William Kramer

Analytic Query Engine **12**
Rovshen Nazarov, Greg Speegle

Lightning Sparks all Around: A Comprehensive Analysis of Popular Distributed Computing Frameworks **19**
Balazs Meszaros, Piyush Harsh, Thomas Michael Bohnert

Adaptive Fusion SQL Engine on Hadoop Framework **27**
Shu-Ming Chang, Chia-Hung Lu, Jiazheng Zhou, Wenguang Chen, Ching-Hsien Hsu, Hung-Chang Hsiao, Yeh-Ching Chung

From In-disk to In-memory Big Data with Hadoop: Performance Experiments with Nucleotide Sequence Data **34**
Atanas Radenski, Louis Ehwerhemuepha, Kyle Anderson

SESSION: BIG DATA, DATA CENTERS AND CLOUD COMPUTING

Improving Backup and Recovery of Modern Data Centers **43**
Zhiwei Zhao, Haolun Yan, Hussain Al-Asaad

On Applying Big Data and Cloud Computing for Quality Improvement in Higher Education **50**
Shakeel Ahmed, Hemant Kumar Mehta

SESSION: BIG DATA SEARCH, MINING METHODS, CLUSTERING, KNOWLEDGE EXTRACTION AND ENGINEERING

Randomized Sorting as a Big Data Search Algorithm **57**
Ivan Avramovic, Dana Richards

Keyword Extraction by KNN considering Similarity among Features **64**
Taeho Jo

A New Random Approach to Dimensionality Reduction **69**
Augustine Nsang, Alheri Maikori, Funmilayo Oguntoyinbo, Hubabah Yusuf

Multi-Label Classification for Mining Big Data **75**
Passent Elkafrawy, Amr Mausad, Awad Khalil

Query-Based Dimensionality Reduction Applied to Images **81**
Augustine Nsang, Donatus Edi, Chinyere Ahanonu

SESSION: APPLICATIONS OF BIG DATA + SOFTWARE TOOLS

Who Influences Whom: Content-based Approach for Predicting Influential Users in Social Networks **89**
Khaled Almgren, Jeongkyu Lee

Big Data Analytics on Acquisition Issues of Public Healthcare **96**
Haibo Wang, Chenhui Zhou, Wei Wang, Yaquan Xu

A Research of Rule Discovery Method for Process Object Based on Large Scale Data **103**
Tao Du, Shouning Qu, Yushui Geng, Xingang Wang

SESSION: POSTER PAPERS

Circadian Rhythm of Hospital Death: Difference Between the Intensive Care Unit and General Room **113**
Jung-Ho Park, Yun-Kyeong Cho, Jong-Ha Lee, Yunjung Lee, In-Cheol Kim, Hyoung-Seob Park, Hyuck-Jun Yoon, Hyungseop Kim, Chang-Wook Nam, Seongwook Han, Seung-Ho Hur, Yoon-Nyun Kim

Predicting Sports Outcomes with a Rank-And-Choose Variable Selection Process **115**
Tao Lin, Ricardo Trujillo, Haibo Wang

The Impact of ICT and Big Data on e-Government **117**
Joseph M. Woodside, Shahram Amiri, Brianne Boldrin

A Strategy for Multimedia Data File Retrieval Using Hadoop Framework **119**
Kyung Chang Kim, Jaekyung Lee

Big Data Analysis based Practical Applications in Construction **121**
Yongho Ko, Seungwoo Han

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

Applying Fuzzy C-Means and Artificial Neural Networks into a High-Order Fuzzy Time Series Prediction Model **125**
Mu-Yen Chen, Hsiu-Sen Chiang

A Study on Framework for Efficient Government R&D Program Feasibility Analysis Using the Bayesian Network and Big Data **132**
JaeHyuk Cho

SESSION: ALGORITHMS AND APPLICATIONS FOR BIG DATA

Performance of an Approximate, Multidimensional Range Index for Big Data **137**
Ronald Wroblewski

From Events to Local Dependence Between Random Units in Big Data **144**
Boyan Dimitrov

Improving the Feature Selection for the Development of Linear Model for Discovery of HIV-1 Integrase Inhibitors **150**
Matineh Kashani Moghaddam, Richard Adrian Galvan, Ahmad Reza Hadaegh

Block-matching Twitter Data for Traffic Analysis **155**
Amal Shuqair, Samuel P. Kozaitis