

2017 IEEE International Conference on Big Data (Big Data 2017)

**Boston, Massachusetts, USA
11-14 December 2017**

Pages 1-685



**IEEE Catalog Number: CFP17BGD-POD
ISBN: 978-1-5386-2716-7**

**Copyright © 2017 by the Institute of Electrical and Electronics Engineers, Inc.
All Rights Reserved**

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP17BGD-POD
ISBN (Print-On-Demand):	978-1-5386-2716-7
ISBN (Online):	978-1-5386-2715-0

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
E-mail: curran@proceedings.com
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

TABLE OF CONTENTS

HUMAN-IN-THE-LOOP APPLIED MACHINE LEARNING	1
<i>Carla E. Brodley</i>	
A MORE OPEN EFFICIENT FUTURE FOR AI DEVELOPMENT AND DATA SCIENCE WITH AN INTRODUCTION TO JULIA	2
<i>Alan Edelman</i>	
CONTEXTUAL REINFORCEMENT LEARNING	3
<i>John Langford</i>	
LARGE-SCALE GRAPH REPRESENTATION LEARNING	4
<i>Jure Leskovec</i>	
BEING “BYTES-ORIENTED” IN HPC LEADS TO AN OPEN BIG DATA/AI ECOSYSTEM AND FURTHER ADVANCES INTO THE POST-MOORE ERA	5
<i>Satoshi Matsuoka</i>	
TEXTSCOPE: ENHANCE HUMAN PERCEPTION VIA TEXT MINING	6
<i>Chengxiang Zhai</i>	
COLLECTIVE SUBJECTIVE LOGIC: SCALABLE UNCERTAINTY-BASED OPINION INFERENCE	7
<i>Feng Chen ; Chunpai Wang ; Jin-Hee Cho</i>	
QUALITY-AWARE AGGREGATION & PREDICTIVE ANALYTICS AT THE EDGE	17
<i>Natascha Harth ; Christos Anagnostopoulos</i>	
ROBUST MULTI-LABEL SEMI-SUPERVISED CLASSIFICATION	27
<i>Sheng Li ; Yun Fu</i>	
LIFELONG MULTI-TASK MULTI-VIEW LEARNING USING LATENT SPACES	37
<i>Xiaoli Li ; Sai Nivedita Chandrasekaran ; Jun Huan</i>	
COMPACT MULTI-CLASS BOOSTED TREES	47
<i>Natalia Ponomareva ; Thomas Colthurst ; Gilbert Hendry ; Salem Haykal ; Soroush Radpour</i>	
CONSTRAINT-AWARE DYNAMIC TRUTH DISCOVERY IN BIG DATA SOCIAL MEDIA SENSING	57
<i>Daniel Yue Zhang ; Dong Wang ; Yang Zhang</i>	
STANDARDIZING BIG EARTH DATACUBES	67
<i>Peter Baumann</i>	
ENHANCING DATA QUALITY BY CLEANING INCONSISTENT BIG RDF DATA	74
<i>Salima Benbernou ; Mourad Ouziri</i>	
ITERATIVE MATRIX CORRELATION FOR BISECTION CLUSTERING	80
<i>Byron J. Gao ; Robert Tung ; Yong Yang</i>	
ENTROPIC DETERMINANTS OF MASSIVE MATRICES	88
<i>Diego Granzio ; Stephen Roberts</i>	
BIG ACTIVE LEARNING	94
<i>Er-Chen Huang ; Hsing-Kuo Pao ; Yuh-Jye Lee</i>	
A NOVEL APPROACH TO OPTIMIZATION OF ITERATIVE MACHINE LEARNING ALGORITHMS: OVER HEAP STRUCTURE	102
<i>Hasan Kurban ; Mehmet M. Dalkilic</i>	
MULTI-VIEW GRAPH LEARNING WITH ADAPTIVE LABEL PROPAGATION	110
<i>Sheng Li ; Hongfu Liu ; Zhiqiang Tao ; Yun Fu</i>	
EXPONENTIAL RANDOM GRAPH MODELS WITH BIG NETWORKS: MAXIMUM PSEUDOLIKELIHOOD ESTIMATION AND THE PARAMETRIC BOOTSTRAP	116
<i>Christian S. Schmid ; Bruce A. Desmarais</i>	
AUTOMATED INDUSTRY CLASSIFICATION WITH DEEP LEARNING	122
<i>Sam Wood ; Rohit Muthyala ; Yi Jin ; Yixing Qin ; Nilaj Rukadikar ; Amit Rai ; Hua Gao</i>	
JOINTLY OPTIMIZING TASK GRANULARITY AND CONCURRENCY FOR IN-MEMORY MAPREDUCE FRAMEWORKS	130
<i>Jonghyun Bae ; Hakbeom Jang ; Wenjing Jin ; Jun Heo ; Jaeyoung Jang ; Joo-Young Hwang ; Sangyeum Cho ; Jae W. Lee</i>	
HOW FAST CAN ONE SCALE DOWN A DISTRIBUTED FILE SYSTEM?	141
<i>Nathanaël Cherièrè ; Gabriel Antoniu</i>	

ATM: A DISTRIBUTED, COLLABORATIVE, SCALABLE SYSTEM FOR AUTOMATED MACHINE LEARNING	151
<i>Thomas Swearingen ; Will Drevo ; Bennett Cyphers ; Alfredo Cuesta-Infante ; Arun Ross ; Kalyan Veeramachaneni</i>	
A DECISION TREE BASED APPROACH TOWARDS ADAPTIVE MODELING OF BIG DATA APPLICATIONS	163
<i>Ioannis Giannakopoulos ; Dimitrios Tsoumakos ; Nectarios Koziris</i>	
CHARACTERIZING AND ACCELERATING INDEXING TECHNIQUES ON DISTRIBUTED ORDERED TABLES	173
<i>Shashank Gugnani ; Xiaoyi Lu ; Houliang Qi ; Li Zha ; Dhableswar K. D K Panda</i>	
OOB_CUDNN: ACCOMMODATING CONVOLUTIONAL NEURAL NETWORKS OVER GPU MEMORY CAPACITY	183
<i>Yuki Ito ; Ryo Matsumiya ; Toshio Endo</i>	
A SEMANTICS-AWARE STORAGE FRAMEWORK FOR SCALABLE PROCESSING OF KNOWLEDGE GRAPHS ON HADOOP	193
<i>Hyeongsik Kim ; Padmashree Ravindra ; Kemafor Anyanwu</i>	
ELASTIC MANAGEMENT OF CLOUD APPLICATIONS USING ADAPTIVE REINFORCEMENT LEARNING	203
<i>Konstantinos Lolos ; Ioannis Konstantinou ; Verena Kantere ; Nectarios Koziris</i>	
PERFORMANCE CHARACTERIZATION AND ACCELERATION OF BIG DATA WORKLOADS ON OPENPOWER SYSTEM	213
<i>Xiaoyi Lu ; Haiyang Shi ; Dipti Shankar ; Dhableswar K. D K Panda</i>	
LOW-LATENCY MULTI-THREADED ENSEMBLE LEARNING FOR DYNAMIC BIG DATA STREAMS	223
<i>Diego Marrón ; Eduard Ayguadé ; José R. Herrero ; Jesse Read ; Albert Bifet</i>	
I/O LOAD BALANCING FOR BIG DATA HPC APPLICATIONS	233
<i>Arnab K. Paul ; Arpit Goyal ; Feiyi Wang ; Sarp Oral ; Ali R. Butt ; Michael J. Brim ; Sangeetha B. Srinivasa</i>	
HARPLDA+: OPTIMIZING LATENT DIRICHLET ALLOCATION FOR PARALLEL EFFICIENCY	243
<i>Bo Peng ; Bingjing Zhang ; Langshi Chen ; Mihai Avram ; Robert Henschel ; Craig Stewart ; Shaojuan Zhu ; Emily McCallum ; Lisa Smith ; Tom Zahner ; Jon Omer ; Judy Qiu</i>	
FAST ACCESS TO COLUMNAR, HIERARCHICALLY NESTED DATA VIA CODE TRANSFORMATION	253
<i>Jim Pivarski ; Peter Elmer ; Brian Bockelman ; Zhe Zhang</i>	
SANZU: A DATA SCIENCE BENCHMARK	263
<i>Alex Watson ; Deepigha Shree Vittal Babu ; Suprio Ray</i>	
SCALING UP DATA-PARALLEL ANALYTICS PLATFORMS: LINEAR ALGEBRAIC OPERATION CASES	273
<i>Luna Xu ; Seung-Hwan Lim ; Min Li ; Ali R. Butt ; Ramakrishnan Kannan</i>	
ROBOTMATA: A FRAMEWORK FOR APPROXIMATE PATTERN MATCHING OF BIG DATA ON AN AUTOMATA PROCESSOR	283
<i>Xiaodong Yu ; Kaixi Hou ; Hao Wang ; Wu-Chun Feng</i>	
MAKING CACHES WORK FOR GRAPH ANALYTICS	293
<i>Yunming Zhang ; Vladimir Kiriansky ; Charith Mendis ; Saman Amarasinghe ; Matei Zaharia</i>	
ON THE USABILITY OF HADOOP MAPREDUCE, APACHE SPARK & APACHE FLINK FOR DATA SCIENCE	303
<i>Bilal Akil ; Ying Zhou ; Uwe Röhm</i>	
ENERGY EFFICIENT STOCHASTIC-BASED DEEP SPIKING NEURAL NETWORKS FOR SPARSE DATASETS	311
<i>Mohammed Alawad ; Hong-Jun Yoon ; Georgia Tourassi</i>	
EXTERNAL MEMORY PIPELINING MADE EASY WITH TPIE	319
<i>Lars Arge ; Mathias Rav ; Svend C. Svendsen ; Jakob Truelsen</i>	
COMPRESSED DOMAIN-SPECIFIC DATA PROCESSING AND ANALYSIS	325
<i>Dapeng Dong ; John Herbert</i>	
UNDERSTANDING AND OPTIMIZING THE PERFORMANCE OF DISTRIBUTED MACHINE LEARNING APPLICATIONS ON APACHE SPARK	331
<i>Celestine Dünner ; Thomas Parnell ; Kubilay Atasü ; Manolis Sifalakis ; Haralampos Pozidis</i>	
OPTIMAL REDUCER PLACEMENT TO MINIMIZE DATA TRANSFER IN MAPREDUCE-STYLE PROCESSING	339
<i>Xiao Meng ; Lukasz Golab</i>	

BIG DATA AND HPC COLLOCATION: USING HPC IDLE RESOURCES FOR BIG DATA ANALYTICS	347
<i>Michael Mercier ; David Glesser ; Yiannis Georgiou ; Olivier Richard</i>	
ETRIKS ANALYTICAL ENVIRONMENT: A MODULAR HIGH PERFORMANCE FRAMEWORK FOR MEDICAL DATA ANALYSIS	353
<i>Axel Oehmichen ; Florian Guitton ; Kai Sun ; Jean Grizet ; Thomas Heinis ; Yike Guo</i>	
MULTI-OBJECTIVE OPTIMIZATION OF SCHEDULING DATAFLOWS ON HETEROGENEOUS CLOUD RESOURCES	361
<i>Ilia Pietri ; Yannis Chronis ; Yannis Ioannidis</i>	
NVMD: NON-VOLATILE MEMORY ASSISTED DESIGN FOR ACCELERATING MAPREDUCE AND DAG EXECUTION FRAMEWORKS ON HPC SYSTEMS	369
<i>Md. Wasi-Ur- Rahman ; Nusrat Sharmin Islam ; Xiaoyi Lu ; Dhabaleswar K. D K Panda</i>	
TOWARDS MEMORY AND COMPUTATION EFFICIENT GRAPH PROCESSING ON SPARK	375
<i>Xinhui Tian ; Yuanqing Guo ; Jianfeng Zhan ; Lei Wang</i>	
SANDPIPER: SCALING PROBABILISTIC INFERENCING TO LARGE SCALE GRAPHICAL MODELS	383
<i>Alexander Ulanov ; Manish Marwah ; Mijung Kim ; Roshan Dathathri ; Carlos Zubieta ; Jun Li</i>	
DIONE: PROFILING SPARK APPLICATIONS EXPLOITING GRAPH SIMILARITY	389
<i>Nikos Zacheilas ; Stathis Maroulis ; Vana Kalogeraki</i>	
ON ON-LINE TASK ASSIGNMENT IN SPATIAL CROWDSOURCING	395
<i>Mohammad Asghari ; Cyrus Shahabi</i>	
QUAD: A QUORUM PROTOCOL FOR ADAPTIVE DATA MANAGEMENT IN THE CLOUD	405
<i>Ilir Fetai ; Alexander Stiemer ; Heiko Schuldt</i>	
SEQUENTIAL ALGORITHMS TO SPLIT AND MERGE ULTRA-HIGH RESOLUTION 3D IMAGES	415
<i>Valérie Hayot-Sasson ; Yongping Gao ; Yuhong Yan ; Tristan Glatard</i>	
SPATIOTEMPORAL RANGE PATTERN QUERIES ON LARGE-SCALE CO-MOVEMENT PATTERN DATASETS	425
<i>Shahab Helmi ; Farnoush Banaei-Kashani</i>	
TOWARDS ROBUST MODELS OF FOOD FLOWS AND THEIR ROLE IN INVASIVE SPECIES SPREAD	435
<i>Srinivasan Venkatramanan ; Sichao Wu ; Bowen Shi ; Achla Marathe ; Madhav Marathe ; Stephen Eubank ; Lalit P. Sah ; A. P. Giri ; Luke A. Colavito ; K. S. Nitin ; V. Sridhar ; R. Asokan ; Rangaswamy Muniappan ; G. Norton ; Abhijin Adiga</i>	
A SINGLE-NODE DATASTORE FOR HIGH-VELOCITY MULTIDIMENSIONAL SENSOR DATA	445
<i>Juan A. Colmenares ; Reza Dorrigiv ; Daniel G. Waddington</i>	
MODEL DRIVEN REVERSE ENGINEERING OF NOSQL PROPERTY GRAPH DATABASES: THE CASE OF NEO4J	453
<i>Isabelle Comyn-Wattiau ; Jacky Akoka</i>	
UNIVERSAL DISTANT READING THROUGH METADATA PROXIES WITH ARCHIVESPARK	459
<i>Helge Holzmann ; Vinay Goel ; Emily Novak Gustainis</i>	
BIG DATA TRANSFER OPTIMIZATION BASED ON OFFLINE KNOWLEDGE DISCOVERY AND ADAPTIVE SAMPLING	465
<i>Md S. Q. Zulkar Nine ; Kemal Guner ; Ziyun Huang ; Xiangyu Wang ; Jinhui Xu ; Tevfik Kosar</i>	
A CLOSED-LOOP DEEP LEARNING ARCHITECTURE FOR ROBUST ACTIVITY RECOGNITION USING WEARABLE SENSORS	473
<i>Ramyar Saeedi ; Skyler Norgaard ; Assefaw H. Gebremedhin</i>	
CSTORAGE: AN EFFICIENT CLASSIFICATION-BASED IMAGE STORAGE SYSTEM IN CLOUD DATACENTERS	480
<i>Haiying Shen ; Heng Zhou</i>	
IN-DEPTH EXPLORATION OF SINGLE-SNAPSHOT LOSSY COMPRESSION TECHNIQUES FOR N-BODY SIMULATIONS	486
<i>Dingwen Tao ; Sheng Di ; Zizhong Chen ; Franck Cappello</i>	
RELIABLE FAKE REVIEW DETECTION VIA MODELING TEMPORAL AND BEHAVIORAL PATTERNS	494
<i>Xian Wu ; Yuxiao Dong ; Jun Tao ; Chao Huang ; Nitesh V. Chawla</i>	
EFFICIENT DIVERSIFIED SET MONITORING FOR MOBILE SENSOR STREAM ENVIRONMENTS	500
<i>Masahiro Yokoyama ; Takahiro Hara ; Sanjay K. Madria</i>	

LOW-RANK SINGULAR VALUE THRESHOLDING FOR RECOVERING MISSING AIR QUALITY DATA	508
<i>Yangwen Yu ; James J. Q. Yu ; Victor O. K. Li ; Jacqueline C. K. Lam</i>	
VISUAL ANALYTICS WITH UNPARALLELED VARIETY SCALING FOR BIG EARTH DATA	514
<i>Lina Yu ; Michael L. Rilee ; Yu Pan ; Feiyu Zhu ; Kwo-Sen Kuo ; Hongfeng Yu</i>	
SEMI-SUPERVISED CONVOLUTIONAL NEURAL NETWORKS FOR HUMAN ACTIVITY RECOGNITION	522
<i>Ming Zeng ; Tong Yu ; Xiao Wang ; Le T. Nguyen ; Ole J. Mengshoel ; Ian Lane</i>	
DETECTING UNMETERED TAXI RIDES FROM TRAJECTORY DATA	530
<i>Xibo Zhou ; Ye Ding ; Fengchao Peng ; Qiong Luo ; Lionel M. Ni</i>	
ESTIMATION OF DISTANCE-BASED METRICS FOR VERY LARGE GRAPHS WITH MINHASH SIGNATURES	536
<i>Giambattista Amati ; Simone Angelini ; Giorgio Gambosi ; Gianluca Rossi ; Paola Vocca</i>	
HIGH-PERFORMANCE GEOMETRIC ALGORITHMS FOR SPARSE COMPUTATION IN BIG DATA ANALYTICS	546
<i>Philipp Baumann ; Dorit S. Hochbaum ; Quico Spaen</i>	
ACTIVE LEARNING BASED NEWS VERACITY DETECTION WITH FEATURE WEIGHTING AND DEEP-SHALLOW FUSION	556
<i>Sreyasee Das Bhattacharjee ; Ashit Talukder ; Bala Venkatram Balantrapu</i>	
EXPLOITING VISUAL AND TEXTUAL NEIGHBORHOOD INFORMATION TO IMPROVE IMAGE-TAG RELEVANCE	566
<i>Chandramani Chaudhary ; Poonam Goyal ; Yi-Ping Phoebe Chen</i>	
INVERSE EXTREME LEARNING MACHINE FOR LEARNING WITH LABEL PROPORTIONS	576
<i>Limeng Cui ; Jiawei Zhang ; Zhensong Chen ; Yong Shi ; Philip S. Yu</i>	
E-CLOG: COUNTING EDGE-CENTRIC LOCAL GRAPHLETS	586
<i>Vachik S. Dave ; Nesreen K. Ahmed ; Mohammad Al Hasan</i>	
MULTISTREAM REGRESSION WITH ASYNCHRONOUS CONCEPT DRIFT DETECTION	596
<i>Bo Dong ; Yifan Li ; Yang Gao ; Ahsanul Haque ; Latifur Khan ; Mohammad M. Masud</i>	
BIAS CORRECTION IN CLUSTERING COEFFICIENT ESTIMATION	606
<i>Roohollah Etemadi ; Jianguo Lu</i>	
CLOSED WALK SAMPLER: AN EFFICIENT METHOD FOR ESTIMATING THE SPECTRAL RADIUS OF LARGE GRAPHS	616
<i>Guyue Han ; Harish Sethu</i>	
ONLINE CITY-SCALE HYPER-LOCAL EVENT DETECTION VIA ANALYSIS OF SOCIAL MEDIA AND HUMAN MOBILITY	626
<i>Jun Hu ; Yuxin Wang ; Ping Li</i>	
DRUM: A RHYTHMIC APPROACH TO INTERACTIVE ANALYTICS ON LARGE DATA	636
<i>Jianfeng Jia ; Chen Li ; Michael J. Carey</i>	
DETECTING CHANGES IN STREAMING DATA WITH INFORMATION-THEORETIC WINDOWING	646
<i>Ryoya Kaneko ; Kohei Miyaguchi ; Kenji Yamanishi</i>	
HYBRID ALGORITHMS FOR SUBGRAPH PATTERN QUERIES IN GRAPH DATABASES	656
<i>Foteini Katsarou ; Nikos Ntarmos ; Peter Triantafillou</i>	
DOMAIN-SPECIFIC HIERARCHICAL SUBGRAPH EXTRACTION: A RECOMMENDATION USE CASE	666
<i>Sarasi Lalithsena ; Sujan Perera ; Pavan Kapanipathi ; Amit Sheth</i>	
COEUS: COMMUNITY DETECTION VIA SEED-SET EXPANSION ON GRAPH STREAMS	676
<i>Panagiotis Liakos ; Alexandros Ntoulas ; Alex Delis</i>	
RHEA: ADAPTIVELY SAMPLING AUTHORITATIVE CONTENT FROM SOCIAL ACTIVITY STREAMS	686
<i>Panagiotis Liakos ; Alexandros Ntoulas ; Alex Delis</i>	
TEXT-BASED GEOLOCATION PREDICTION OF SOCIAL MEDIA USERS WITH NEURAL NETWORKS	696
<i>Ismeni Lourentzou ; Alex Morales ; Chengxiang Zhai</i>	
CRACK RANDOM FOREST FOR ARBITRARY LARGE DATASETS	706
<i>Alessandro Lulli ; Luca Oneto ; Davide Anguita</i>	
S-ISOMAP++: MULTI MANIFOLD LEARNING FROM STREAMING DATA	716
<i>Suchismit Mahapatra ; Varun Chandola</i>	
A SCALABLE MODEL FOR TRACKING TOPICAL EVOLUTION IN LARGE DOCUMENT COLLECTIONS	726
<i>Sheikh Motahar Naim ; Arnold P. Boedihardjo ; M. Shahriar Hossain</i>	

ERROR-ROBUST MULTI-VIEW CLUSTERING	736
<i>Mehrmaz Najafi ; Lifang He ; Philip S. Yu</i>	
HOLISTIC AND SCALABLE RANKING OF RDF DATA	746
<i>Axel-Cyrille ; Ngonga Ngomo ; Michael Hoffmann ; Ricardo Usbeck ; Kunal Jha</i>	
A COMPARATIVE STUDY OF MATRIX FACTORIZATION AND RANDOM WALK WITH RESTART IN RECOMMENDER SYSTEMS	756
<i>Haekyu Park ; Jinhong Jung ; U. Kang</i>	
VIGAN: MISSING VIEW IMPUTATION WITH GENERATIVE ADVERSARIAL NETWORKS	766
<i>Chao Shang ; Aaron Palmer ; Jiangwen Sun ; Ko-Shin Chen ; Jin Lu ; Jinbo Bi</i>	
TIERED SAMPLING: AN EFFICIENT METHOD FOR APPROXIMATE COUNTING SPARSE MOTIFS IN MASSIVE GRAPH STREAMS	776
<i>Lorenzo De Stefani ; Erisa Terolli ; Eli Upfal</i>	
A FAST NON-VOLATILE MEMORY AWARE ALGORITHM FOR GENERATING RANDOM SCALE-FREE NETWORKS	787
<i>Cheng-Chin Tu ; Mi-Yen Yeh ; Tei-Wei Kuo</i>	
MRATTRACTOR: DETECTING COMMUNITIES FROM LARGE-SCALE GRAPHS	797
<i>Nguyen Vo ; Kyumin Lee ; Thanh Tran</i>	
POTENTIALITY OF HEALTHCARE BIG DATA: IMPROVING SEARCH BY AUTOMATIC QUERY REFORMULATION	807
<i>Yueyao Wang ; Qimmin Hu ; Yang Song ; Liang He</i>	
SAMPLING ALGORITHMS TO UPDATE TRUNCATED SVD	817
<i>Ichitaro Yamazaki ; Stanimire Tomov ; Jack Dongarra</i>	
DISTRIBUTED TOP-N LOCAL OUTLIER DETECTION IN BIG DATA	827
<i>Yizhou Yan ; Lei Cao ; Elke A. Rundensteiner</i>	
RECTANGULAR HASH TABLE: BLOOM FILTER AND BITMAP ASSISTED HASH TABLE WITH HIGH SPEED	837
<i>Tong Yang ; Binchao Yin ; Hang Li ; Muhammad Shahzad ; Steve Uhlig ; Bin Cm ; Xiaoming Li</i>	
LARGE-SCALE JOINT TOPIC, SENTIMENT & USER PREFERENCE ANALYSIS FOR ONLINE REVIEWS	847
<i>Xinli Yu ; Zheng Chen ; Wei-Shih Yang ; Xiaohua Hu ; Erjia Yan ; Guangrong Li</i>	
IMWALKMF: JOINT MATRIX FACTORIZATION AND IMPLICIT WALK INTEGRATIVE LEARNING FOR RECOMMENDATION	857
<i>Chuxu Zhang ; Lu Yu ; Xiangliang Zhang ; Nitesh Chawla</i>	
HIERARCHICAL COLLABORATIVE EMBEDDING FOR CONTEXT-AWARE RECOMMENDATIONS	867
<i>Lei Zheng ; Bokai Cao ; Vahid Noroozi ; Philip S. Yu ; Nianzu Ma</i>	
MINING PROS AND CONS OF ACTIONS FROM SOCIAL MEDIA FOR DECISION SUPPORT	877
<i>Ebad Ahmadzadeh ; Philip K. Chan</i>	
DISTRIBUTED BAYESIAN PIECEWISE SPARSE LINEAR MODELS	883
<i>Masato Asahara ; Ryohei Fujimaki</i>	
LINEAR-COMPLEXITY RELAXED WORD MOVER'S DISTANCE WITH GPU ACCELERATION	889
<i>Kubilay Atasü ; Thomas Parnell ; Celestine Dünner ; Manolis Sifalakis ; Haralampos Pozidis ; Vasileios Vasileiadis ; Michail Vlachos ; Cesar Berrospi ; Abdel Labbi</i>	
QUALITY-EFFICIENCY TRADE-OFFS IN MACHINE LEARNING FOR TEXT PROCESSING	897
<i>Ricardo Baeza-Yates ; Zeinab Liaghat</i>	
FAST GRAPH SCAN STATISTICS OPTIMIZATION USING ALGEBRAIC FINGERPRINTS	905
<i>Jose Cadena ; Saliya Ekanayake ; Anil Vullikanti</i>	
A DISTRIBUTED ROUGH SET THEORY BASED ALGORITHM FOR AN EFFICIENT BIG DATA PRE-PROCESSING UNDER THE SPARK FRAMEWORK	911
<i>Zaineb Chelly Dagdia ; Christine Zarges ; Gaël Beck ; Mustapha Lebbah</i>	
JUDICIOUS SETTING OF DYNAMIC TIME WARPING'S WINDOW WIDTH ALLOWS MORE ACCURATE CLASSIFICATION OF TIME SERIES	917
<i>Hoang Anh Dau ; Diego Furtado Silva ; François Petitjean ; Germain Forestier ; Anthony Bagnall ; Eamonn Keogh</i>	
TOWARD GRANULAR KNOWLEDGE ANALYTICS FOR DATA INTELLIGENCE: EXTRACTING GRANULAR ENTITY-RELATIONSHIP GRAPHS FOR KNOWLEDGE PROFILING	923
<i>Alexander Denzler ; Michael Kaufmann</i>	
DISTRIBUTED DECISION TREE V.2.0	929
<i>Ankit Desai ; Sanjay Chaudhary</i>	

AN OPEN-SOURCE TOOL FOR THE TRANSCRIPTION OF PAPER-SPREADSHEET DATA: CODE AND SUPPLEMENTAL MATERIALS AVAILABLE ONLINE: HTTPS://GITHUB.COM/DESKOOL/IMAGES TO SPREADSHEETS	935
<i>Mohammad M. Ghassemi ; Willow Jarvis ; Tuka Alhanai ; Emery N. Brown ; Roger G. Mark ; M. Brandon Westover</i>	
ANYFI: AN ANYTIME FREQUENT ITEMSET MINING ALGORITHM FOR DATA STREAMS	942
<i>Poonam Goyal ; Jagat Sesh Challa ; Shivin Shrivastava ; Navneet Goyal</i>	
DISCOVERING POTENTIAL TRAFFIC RISKS IN JAPAN USING A SUPERVISED LEARNING APPROACH	948
<i>Tatsuru Kobayashi ; Shin Matsushima ; Taito Lee ; Kenji Yamanishi</i>	
DATA CONTEXT INFORMED DATA WRANGLING	956
<i>Martin Koehler ; Alex Bogatu ; Cristina Civili ; Nikolaos Konstantinou ; Edward Abel ; Alvaro A. A. Fernandes ; John Keane ; Leonid Libkin ; Norman W. Paton</i>	
FISHING IN THE STREAM: SIMILARITY SEARCH OVER ENDLESS DATA	964
<i>Naama Kraus ; David Carmel ; Idit Keidar</i>	
GRAPHICAL APPROACH FOR INFLUENCE MAXIMIZATION IN SOCIAL NETWORKS UNDER GENERIC THRESHOLD-BASED NON-SUBMODULAR MODEL	970
<i>Liang Ma ; Guohong Cao ; Lance Kaplan</i>	
A DISTRIBUTED K-CORE DECOMPOSITION ALGORITHM ON SPARK	976
<i>Aritra Mandal ; Mohammad Al Hasan</i>	
EVENT PATTERN DISCOVERY BY KEYWORDS IN GRAPH STREAMS	982
<i>Mohammad Hossein Namaki ; Peng Lin ; Yinghui Wu</i>	
QUERYABLE COMPRESSION ON STREAMING SOCIAL NETWORKS	988
<i>Michael Nelson ; Sridhar Radhakrishnan ; Amlan Chatterjee ; Chandra N. Sekharan</i>	
EVENT-BASED NON-PARAMETRIC CLUSTERING OF TEAM SPORT TRAJECTORIES	994
<i>Fengchao Peng ; Yudian Ji ; Qiong Luo ; Lionel M. Ni</i>	
APPLICATION-SPECIFIC GRAPH SAMPLING FOR FREQUENT SUBGRAPH MINING AND COMMUNITY DETECTION	1000
<i>Sumit Purohit ; Sutanay Choudhury ; Lawrence B. Holder</i>	
DISCOVERING CO-OCCURRENCE PATTERNS OF HETEROGENEOUS EVENTS FROM UNEVENLY-DISTRIBUTED SPATIOTEMPORAL DATA	1006
<i>Hung Tran-The ; Koji Zettsu</i>	
MICRO-CLUSTERING BY DATA POLISHING	1012
<i>Takeaki Uno ; Hiroki Maegawa ; Takanoobu Nakahara ; Yukinobu Hamuro ; Ryo Yoshinaka ; Makoto Tatsuta</i>	
BRINGING SEMANTIC STRUCTURES TO USER INTENT DETECTION IN ONLINE MEDICAL QUERIES	1019
<i>Chenwei Zhang ; Nan Du ; Wei Fan ; Yaliang Li ; Chun-Ta Lu ; Philip S. Yu</i>	
LARGE-SCALE POINT-OF-INTEREST CATEGORY PREDICTION USING NATURAL LANGUAGE PROCESSING MODELS	1027
<i>Daniel Yue Zhang ; Dong Wang ; Hao Zheng ; Xin Mu ; Qi Li ; Yang Zhang</i>	
SHADE: A DIFFERENTIALLY-PRIVATE WRAPPER FOR ENTERPRISE BIG DATA	1033
<i>Alexander Heifetz ; Vaikkunth Mughunthan ; Lalana Kagal</i>	
GROUP PRIVACY-AWARE DISCLOSURE OF ASSOCIATION GRAPH DATA	1043
<i>Balaji Palanisamy ; Chao Li ; Prashant Krishnamurthy</i>	
CONTAMINANT REMOVAL FOR ANDROID MALWARE DETECTION SYSTEMS	1053
<i>Lichao Sun ; Xiaokai Wei ; Jiawei Zhang ; Lifang He ; Philip S. Yu ; Witawas Srisa-An</i>	
BOOSTING THE PHISHING DETECTION PERFORMANCE BY SEMANTIC ANALYSIS	1063
<i>Xi Zhang ; Yu Zeng ; Xiao-Bo Jin ; Zhi-Wei Yan ; Guang-Gang Geng</i>	
SETTING THE THRESHOLD FOR HIGH THROUGHPUT DETECTORS: A MATHEMATICAL APPROACH FOR ENSEMBLES OF DYNAMIC, HETEROGENEOUS, PROBABILISTIC ANOMALY DETECTORS	1071
<i>Robert A. Bridges ; Jessie D. Jamieson ; Joel W. Reed</i>	
WEATHERMAN: EXPOSING WEATHER-BASED PRIVACY THREATS IN BIG ENERGY DATA	1079
<i>Dong Chen ; David Irwin</i>	
DISCRIMINATION DETECTION BY CAUSAL EFFECT ESTIMATION	1087
<i>Jiuyong Li ; Jixue Liu ; Lin Liu ; Thuc Duy Le ; Saisai Ma ; Yizhao Han</i>	
WEAC: WORD EMBEDDINGS FOR ANOMALY CLASSIFICATION FROM EVENT LOGS	1095
<i>Amit Pande ; Vishal Ahuja</i>	
PRIVACY-PROTECTED PLACE OF ACTIVITY MINING ON BIG LOCATION DATA	1101
<i>Shuo Wang ; Richard Sinnott ; Surya Nepal</i>	

SENSITIVE GAZETTEER DISCOVERY AND PROTECTION FOR MOBILE SOCIAL MEDIA USERS	1109
<i>Shuo Wang ; Richard Sinnott ; Surya Nepal</i>	
DIFFERENTIALLY PRIVATE QUERY LEARNING: FROM DATA PUBLISHING TO MODEL PUBLISHING	1117
<i>Tianqing Zhu ; Ping Xiong ; Gang Li ; Wanlei Zhou ; Philip S. Yu</i>	
THE ML TEST SCORE: A RUBRIC FOR ML PRODUCTION READINESS AND TECHNICAL DEBT REDUCTION	1123
<i>Eric Breck ; Shanqing Cai ; Eric Nielsen ; Michael Salib ; D. Sculley</i>	
BTCT: A NEW FRAMEWORK FOR IDENTIFYING CONGESTION CASCADES USING BUS TRAJECTORY DATA	1133
<i>Meng-Fen Chiang ; Ee-Peng Lim ; Wang-Chien Lee ; Agus Trisnajaya Kwee</i>	
APPLICATION OF BIG DATA ANALYTICS IN PROCESS SAFETY AND RISK MANAGEMENT	1143
<i>Pankaj Goel ; Aniruddha Datta ; M. Sam Mannan</i>	
ENABLING VERSATILE ANALYSIS OF LARGE SCALE TRAFFIC VIDEO DATA WITH DEEP LEARNING AND HIVEQL	1153
<i>Lei Huang ; Weijia Xu ; Si Liu ; Venkatesh Pandey ; Natalia Ruiz Juri</i>	
FAST INTERPOLATION OF GRID DATA AT A NON-GRID POINT	1163
<i>Hiroshi Inoue</i>	
JOINT SPARSE AUTO-ENCODER: A SEMI-SUPERVISED SPATIO-TEMPORAL APPROACH IN MAPPING LARGE-SCALE CROPLANDS	1173
<i>Xiaowei Jia ; Yifan Hu ; Ankush Khandelwal ; Anuj Karpatne ; Vipin Kumar</i>	
MULTI-STEP PREDICTION WITH MISSING SMART SENSOR DATA USING MULTI-TASK GAUSSIAN PROCESSES	1183
<i>Pasan Karunaratne ; Masud Moshtaghi ; Shanika Karunasekera ; Aaron Harwood ; Trevor Cohn</i>	
BAYESIAN MULTI-VIEW MODELS FOR MEMBER-JOB MATCHING AND PERSONALIZED SKILL RECOMMENDATIONS	1193
<i>Abhinav Maurya ; Rahul Telang</i>	
AUTOMATED SCALABLE DETECTION OF LOCATION-SPECIFIC SANTA ANA CONDITIONS FROM WEATHER DATA USING UNSUPERVISED LEARNING	1203
<i>Mai H. Nguyen ; Daniel Crawl ; Jiaxin Li ; Dylan Uys ; Ilkay Altintas</i>	
HEALTHEDGE: TASK SCHEDULING FOR EDGE COMPUTING WITH HEALTH EMERGENCY AND HUMAN BEHAVIOR CONSIDERATION IN SMART HOMES	1213
<i>Haoyu Wang ; Jiaqi Gong ; Yan Zhuang ; Haiying Shen ; John Lach</i>	
CONNECTING EMERGING RELATIONSHIPS FROM NEWS VIA TENSOR FACTORIZATION	1223
<i>Jingyuan Zhang ; Chun-Ta Lu ; Bokai Cao ; Yi Chang ; Philip S. Yu</i>	
LSTM FOR SEPTIC SHOCK: ADDING UNRELIABLE LABELS TO RELIABLE PREDICTIONS	1233
<i>Yuan Zhang ; Chen Lin ; Min Chi ; Julie Ivy ; Muge Capan ; Jeanne M. Huddleston</i>	
A DATA-DRIVEN CONGESTION DIFFUSION MODEL FOR CHARACTERIZING TRAFFIC IN METRO CITY SCALES	1243
<i>Baoxin Zhao ; Chengzhong Xu ; Siyuan Liu</i>	
ANALYSIS OF THE TERM 'BIG DATA': USAGE IN BIOMEDICAL PUBLICATIONS	1253
<i>A. J. Van Altena ; P. D. Moerland ; A. H. Zwinderman ; S. D. Olabarriaga</i>	
PREDICTING TREATMENT REPETITIONS IN THE IMPLANT DENTURE THERAPY PROCESS	1259
<i>Marzieh Bakhshandeh ; Dennis M. M. Schunselaar ; Henrik Leopold ; Hajo A. Reijers</i>	
PERSONALIZED FLIGHT RECOMMENDATIONS VIA PAIRED CHOICE MODELING	1265
<i>Jian Cao ; Fangzhou Yang ; Yuchang Xu ; Yudong Tan ; Quanwu Xiao</i>	
SEQ2IMG: A SEQUENCE-TO-IMAGE BASED APPROACH TOWARDS IP TRAFFIC CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORKS	1271
<i>Zhitang Chen ; Ke He ; Jian Li ; Yanhui Geng</i>	
OREONET: DEEP CONVOLUTIONAL NETWORK FOR OIL RESERVOIR OPTIMIZATION	1277
<i>Chung Ming Cheung ; Palash Goyal ; Viktor K. Prasanna ; Arash Saber Tehrani</i>	
A DATA-DRIVEN APPROACH TO PREDICT NOX-EMISSIONS OF GAS TURBINES	1283
<i>Giuseppe Cuccu ; Somayeh Danafar ; Philippe Cudré-Mauroux ; Martin Gassner ; Stefano Bernero ; Krzysztof Kryszczuk</i>	
TWO-LEVEL CLUSTERING FAST BETWEENNESS CENTRALITY COMPUTATION FOR REQUIREMENT-DRIVEN APPROXIMATION	1289
<i>Angelo Furno ; N. E. El Faouzi ; Rajesh Sharma ; Eugenio Zimeo</i>	
CELLULAR NETWORK CONFIGURATION VIA ONLINE LEARNING AND JOINT OPTIMIZATION	1295
<i>Xueying Guo ; George Trimponias ; Xiaoxiao Wang ; Zhitang Chen ; Yanhui Geng ; Xin Liu</i>	

T-BMIRT: ESTIMATING REPRESENTATIONS OF STUDENT KNOWLEDGE AND EDUCATIONAL COMPONENTS IN ONLINE EDUCATION	1301
<i>Jiankun Huang ; Wenjun Wu</i>	
FORECASTING THE RISE AND FALL OF VOLATILE POINT-OF-INTERESTS	1307
<i>Xinjiang Lu ; Zhiwen Yu ; Chuanren Liu ; Yanchi Liu ; Hui Xiong ; Bin Guo</i>	
PREDICTING REGIONAL ECONOMIC INDICES USING BIG DATA OF INDIVIDUAL BANK CARD TRANSACTIONS	1313
<i>Stanislav Sobolevsky ; Emanuele Massaro ; Iva Bojic ; Juan Murillo Arias ; Carlo Ratti</i>	
TRAVEL PURPOSE INFERENCE WITH GPS TRAJECTORIES, POIS, AND GEO-TAGGED SOCIAL MEDIA DATA	1319
<i>Chuisheng Meng ; Yu Cui ; Qing He ; Lu Su ; Jing Gao</i>	
DISCOVERING SCIENTIFIC INFLUENCE USING CROSS-DOMAIN DYNAMIC TOPIC MODELING	1325
<i>Jennifer Sleeman ; Milton Halem ; Tim Finin ; Mark Cane</i>	
REPAIR: RECOMMEND POLITICAL ACTORS IN REAL-TIME FROM NEWS WEBSITES	1333
<i>Mohiuddin Solaimani ; Sayeed Salam ; Latifur Khan ; Patrick T. Brandt ; Vito D'Orazio</i>	
PERSONALIZED TRAVEL MODE DETECTION WITH SMARTPHONE SENSORS	1341
<i>Xing Su ; Yuan Yao ; Qing He ; Jie Lu ; Hanghang Tong</i>	
A COMPARATIVE ANALYSIS OF STATE-OF-THE-ART SQL-ON-HADOOP SYSTEMS FOR INTERACTIVE ANALYTICS	1349
<i>Ashish Tapdiya ; Daniel Fabbri</i>	
IDENTIFYING AND QUANTIFYING NONLINEAR STRUCTURED RELATIONSHIPS IN COMPLEX MANUFACTURAL SYSTEMS	1357
<i>Tingyang Xu ; Tan Yan ; Dongjin Song ; Wei Cheng ; Haifeng Chen ; Geoff Jiang ; Jinbo Bi</i>	
OTPS: A DECISION SUPPORT SERVICE FOR OPTIMAL AIRFARE TICKET PURCHASE	1363
<i>Yuchang Xu ; Jian Cao</i>	
PRODUCT FUNCTION NEED RECOGNITION VIA SEMI-SUPERVISED ATTENTION NETWORK	1369
<i>Hu Xu ; Sihong Xie ; Lei Shu ; Philip S. Yu</i>	
EXPLORING THE DYNAMICS OF SURGE PRICING IN MOBILITY-ON-DEMAND TAXI SERVICES	1375
<i>Wenbo Zhang ; Dheeraj Kumar ; Satish V. Ukkusuri</i>	
APPLICATION OF DYNAMIC LOGISTIC REGRESSION WITH UNSCENTED KALMAN FILTER IN PREDICTIVE CODING	1381
<i>Yihua Shi Astle ; Xuning Tang ; Craig Freeman</i>	
RAVEN: WEB-BASED SMART HOME EXPLORATION SYSTEM THROUGH INTERACTIVE PATTERN DISCOVERY	1390
<i>Mansurul A Bhuiyan ; Mohammad Al Hasan</i>	
IMPLEMENTING SCALABLE STRUCTURED MACHINE LEARNING FOR BIG DATA IN THE SAKE PROJECT	1400
<i>Simon Bin ; Patrick Westphal ; Jens Lehmann ; Axel Ngonga</i>	
FAST BOTNET DETECTION FROM STREAMING LOGS USING ONLINE LANZOS METHOD	1408
<i>Zheng Chen ; Xinli Yu ; Chi Zhang ; Jin Zhang ; Cui Lin ; Bo Song ; Jianliang Gao ; Xiaohua Hu ; Wei-Shih Yang ; Erjia Yan</i>	
REPRESENTATIVENESS OF LATENT DIRICHLET ALLOCATION TOPICS ESTIMATED FROM DATA SAMPLES WITH APPLICATION TO COMMON CRAWL	1418
<i>Yuheng Du ; Alexander Herzog ; Andre Luckow ; Ramu Nerella ; Christopher Gropp ; Amy Apon</i>	
EMPIRICAL EVALUATIONS OF ACTIVE LEARNING STRATEGIES IN LEGAL DOCUMENT REVIEW	1428
<i>Rishi Chhatwal ; Nathaniel Huber-Fliflet ; Robert Keeling ; Jianping Zhang ; Haozhen Zhao</i>	
TOPIC MODELS FOR RFID DATA MODELING AND LOCALIZATION	1438
<i>T. F. Kennedy ; Robert S. Provence ; James L. Broyan ; Patrick W. Fink ; Phong H. Ngo ; Lazaro D. Rodriguez</i>	
WHAT IS SKIPPED: FINDING DESIRABLE ITEMS IN E-COMMERCE SEARCH BY DISCOVERING THE WORST TITLE TOKENS	1447
<i>Ishita K Khan ; Prathyusha Senthil Kumar ; Daniel Miranda ; David Goldberg</i>	
RANKING THE IMPORTANCE OF ONTOLOGY CONCEPTS USING DOCUMENT SUMMARIZATION TECHNIQUES	1457
<i>Youngho Kim ; Petros Zerfos ; Vadim Sheinin ; Nancy Greco</i>	
PERFORMANCE OPTIMIZATION IN SCALE-OUT STORAGE USING DESIGN OF EXPERIMENT AS HEURISTIC	1467
<i>Lay Wai Kong</i>	

A STUDY ON INTELLIGENT PERSONALIZED PUSH NOTIFICATION WITH USER HISTORY	1475
<i>Hyunjong Lee ; Youngin Jo ; Sanghyuk Chun ; Kwangseob Kim</i>	
REUTERS TRACER: TOWARD AUTOMATED NEWS PRODUCTION USING LARGE SCALE SOCIAL MEDIA DATA	1483
<i>Xiaomo Liu ; Armineh Nourbakhsh ; Quanzhi Li ; Sameena Shah ; Robert Martin ; John Duprey</i>	
INTEGRATED ACCESS TO BIG DATA POLYSTORES THROUGH A KNOWLEDGE-DRIVEN FRAMEWORK	1494
<i>Justin McHugh ; Paul E. Cuddihy ; Jenny Weisenberg Williams ; Kareem S. Aggour ; Vijay S. Kumar ; Varish Mulwad</i>	
PREDICTING OVER-INDEBTEDNESS ON BATCH AND STREAMING DATA	1504
<i>Jacob Montiel ; Albert Bifet ; Talel Abdessalem</i>	
APP-SOFT: APPLICATION CHARACTERISTICS-DRIVEN SON TO OPTIMIZE 4G/5G NETWORK PERFORMANCE AND QUALITY OF EXPERIENCE	1514
<i>Ye Ouyang ; Zhongyuan Li ; Le Su ; Wenyuan Lu ; Zhenyi Lin</i>	
A CONFIGURABLE, BIG DATA SYSTEM FOR ON-DEMAND HEALTHCARE COST PREDICTION	1524
<i>Karthikeyan Natesan Ramamurthy ; Dennis Wei ; Emily Ray ; Moninder Singh ; Vijay Iyengar ; Dmitriy Katz-Rogozhnikov ; Jingwei Yang ; Kevin N. Tran ; Gigi Yuen-Reed</i>	
DEPENDENCY ANALYSIS OF CLOUD APPLICATIONS FOR PERFORMANCE MONITORING USING RECURRENT NEURAL NETWORKS	1534
<i>Syed Yousaf Shah ; Zengwen Yuan ; Songwu Lu ; Petros Zerfos</i>	
HELP ME FIND A JOB: A GRAPH-BASED APPROACH FOR JOB RECOMMENDATION AT SCALE	1544
<i>Walid Shalaby ; Bahaaeddin Alaila ; Mohammed Korayem ; Layla Pournajaf ; Khalifeh Aljadda ; Shannon Quinn ; Wlodek Zadrozny</i>	
FLUX: Groupon's Automated, Scalable, Extensible Machine Learning Platform	1554
<i>Derrick C. Spell ; Xiao-Han T. Zeng ; Jae Young Chung ; Bahador Nooraei ; Richard T. Shomer ; Ling-Yong Wang ; James C. Gibson ; Daniel Kirsche</i>	
A DATA-DRIVEN APPROACH FOR MULTIVARIATE CONTEXTUALIZED ANOMALY DETECTION: INDUSTRY USE CASE	1560
<i>Nenad Stojanovic ; Marko Dinic ; Ljiljana Stojanovic</i>	
A COGNITIVE ASSISTANT FOR RISK IDENTIFICATION AND MODELING	1570
<i>Dharmashankar Subramanian ; Debarun Bhattacharjya ; Ruben R. Torrado ; Jeff Kephart ; Vijil Chenthamarakshan ; Jesus Rios</i>	
SCALABLE TIME-VERSIONING SUPPORT FOR PROPERTY GRAPH DATABASES	1580
<i>Warut D. Vijitbenjaronk ; Jinho Lee ; Toyotaro Suzumura ; Gabriel Tanase</i>	
TRENDI: TRACKING STORIES IN NEWS AND MICROBLOGS VIA EMERGING, EVOLVING AND FADING TOPICS	1590
<i>Xuchao Zhang ; Liang Zhao ; Zhiqian Chen ; Arnold P. Boedihardjo ; Jing Dai ; Chang-Tien Lu</i>	
SMART: SPONSORED MOBILE APP RECOMMENDATION BY BALANCING APP DOWNLOADS AND APPSTORE PROFIT	1600
<i>Zhiwei Zhang ; Ning Chen ; Jun Wang ; Luo Si</i>	
A GAMMA-BASED REGRESSION FOR WINNING PRICE ESTIMATION IN REAL-TIME BIDDING ADVERTISING	1610
<i>Wen-Yuan Zhu ; Wen-Yueh Shih ; Ying-Hsuan Lee ; Wen-Chih Peng ; Jiun-Long Huang</i>	
DEMYSTIFYING DARK MATTER FOR ONLINE EXPERIMENTATION	1620
<i>Nirupama Appikala ; Miao Chen ; Michael Natkovich ; Joshua Walters</i>	
DETECTING AND SUMMARIZING EMERGENT EVENTS IN MICROBLOGS AND SOCIAL MEDIA STREAMS BY DYNAMIC CENTRALITIES	1627
<i>Neela Avudaiappan ; Alexander Herzog ; Sneha Kadam ; Yuheng Du ; Jason Thatche ; Ilya Safro</i>	
FASTER ONLINE EXPERIMENTATION BY ELIMINATING TRADITIONAL A/A VALIDATION	1635
<i>Russell Chen ; Miao Chen ; Mahendrasinh Ramsinh Jadav ; Joonsuk Bae ; Don Matheson</i>	
BBC: A DSL FOR DESIGNING CLOUD-BASED HETEROGENEOUS BIGDATA PIPELINES	1642
<i>Ferosh Jacob ; Ilamgumaran Karunanithi ; Pramod Salian ; Ravi Sambhu</i>	
ARCHITECTURAL CONSIDERATIONS FOR HIGHLY SCALABLE COMPUTING TO SUPPORT ON-DEMAND VIDEO ANALYTICS	1646
<i>George Mathew</i>	
SCALABLE DISTRIBUTED CHANGE DETECTION AND ITS APPLICATION TO MARITIME TRAFFIC	1650
<i>Leonardo M. Millefiori ; Paolo Braca ; Gianfranco Arcieri</i>	

CONNECTED HEALTH: OPPORTUNITIES AND CHALLENGES	1658
<i>Ankita R. Nambiar ; Nikitha Reddy ; Debojyoti Dutta</i>	
PREDICTIVE EDGE COMPUTING FOR TIME SERIES OF INDUSTRIAL IOT AND LARGE SCALE CRITICAL INFRASTRUCTURE BASED ON OPEN-SOURCE SOFTWARE ANALYTIC OF BIG DATA	1663
<i>Emmanuel Oyekanlu</i>	
LINKING MANY UNUSUAL CO-INCIDENCES	1670
<i>Kevin B. Pratt</i>	
ON EVENT-DRIVEN KNOWLEDGE GRAPH COMPLETION IN DIGITAL FACTORIES	1676
<i>Martin Ringsquandl ; Evgeny Kharlamov ; Daria Stepanova ; Steffen Lamparter ; Raffaello Lepratti ; Ian Horrocks ; Peer Kröger</i>	
KNOWLEDGE EXTRACTION FROM MARITIME SPATIOTEMPORAL DATA: AN EVALUATION OF CLUSTERING ALGORITHMS ON BIG DATA	1682
<i>Giannis Spiliopoulos ; Konstantinos Chatzikokolakis ; Dimitrios Zissis ; Evmorfia Biliri ; Dimitrios Papaspyros ; Giannis Tsapelas ; Spyros Mouzakitis</i>	
TRACES: GENERATING TWITTER STORIES VIA SHARED SUBSPACE AND TEMPORAL SMOOTHNESS	1688
<i>Xuchao Zhang ; Zhiqian Chen ; Liang Zhao ; Arnold P. Boedihardjo ; Chang-Tien Lu</i>	
TRACKING AND PREDICTING THE EVOLUTION OF RESEARCH TOPICS IN SCIENTIFIC LITERATURE	1694
<i>Christine Balili ; Aviv Segev ; Uichin Lee</i>	
TOWARDS A SEMANTIC KEYWORD SEARCH OVER INDUSTRIAL KNOWLEDGE GRAPHS (EXTENDED ABSTRACT)	1698
<i>Gong Cheng ; Evgeny Kharlamov</i>	
DESIGNING A HIGH PERFORMANCE CLUSTER FOR LARGE-SCALE SQL-ON-HADOOP ANALYTICS	1701
<i>Ajay Dholakia ; Prasad Venkatachar ; Kshitij Doshi ; Ravikanth Durgavajhala ; Stewart Tate ; Berni Schiefer ; Matthew Sheard ; Ramnath Sai Sagar</i>	
REAL TIME SEMANTIC ENRICHMENT OF BROADCAST CONTENT IN THE BIG DATA AGE	1704
<i>Maurizio Montagnuolo ; Alberto Messina ; Nicolò Bidotti ; Paolo Platter ; Alessio Bosca</i>	
ON THE IMPROVEMENT OF CLASSIFYING EEG RECORDINGS USING NEURAL NETWORKS	1709
<i>Yiran Zhao ; Shuochao Yao ; Shaohan Hu ; Shiyu Chang ; Raghu Ganti ; Mudhakar Srivatsa ; Shen Li ; Tarek Abdelzaher</i>	
A ROBUST INTERNET ABUSE DETECTION METHOD	1712
<i>Zhou Fa ; Guang-Gang Geng ; Zhi-Wei Yan ; Xiao-Dong Lee</i>	
MANUFACTURING AND CONTRACT SERVICE NETWORKS: COMPOSITION, OPTIMIZATION AND TRADEOFF ANALYSIS BASED ON A REUSABLE REPOSITORY OF PERFORMANCE MODELS	1716
<i>Alexander Brodsky ; Mohan Krishnamoorthy ; M. Omar Nachawati ; William Z. Bernstein ; Daniel A. Menascé</i>	
AUTOMATIC LOCALIZATION OF CASTING DEFECTS WITH CONVOLUTIONAL NEURAL NETWORKS	1726
<i>Max Ferguson ; Ronay Ak ; Yung-Tsun Tina Lee ; Kincho H. Law</i>	
A DATA-DRIVEN APPROACH FOR IMPROVING SUSTAINABILITY ASSESSMENT IN ADVANCED MANUFACTURING	1736
<i>Yunpeng Li ; Heng Zhang ; Utpal Roy ; Y. Tina Lee</i>	
ISSUES IN SYNTHETIC DATA GENERATION FOR ADVANCED MANUFACTURING	1746
<i>Don Libes ; David Lechevalier ; Sanjay Jain</i>	
ESTIMATION OF ONLINE TOOL WEAR IN TURNING PROCESSES USING RECURRENCE QUANTIFICATION ANALYSIS (RQA)	1755
<i>Srinivasan Radhakrishnan ; Yung-Tsun Tina Lee ; Sagar Kamarathi</i>	
STATISTICALLY-SUBSTANTIATED DENSITY CHARACTERIZATIONS OF ADDITIVELY MANUFACTURED STEEL ALLOYS THROUGH VERIFICATION, VALIDATION, AND UNCERTAINTY QUANTIFICATION	1760
<i>Heather M. Reed ; Richard P. Vinci ; Corbin Robeck ; Trevor Verdonik ; Michael Pires ; Maria Castro ; Wojciech Z. Misiolek ; Christina Viau Haden</i>	
HYBRID DATAFICATION OF MAINTENANCE LOGS FROM AI-ASSISTED HUMAN TAGS	1769
<i>Thurston Sexton ; Michael P. Brundage ; Michael Hoffman ; K C Morris</i>	
DATA TREATMENT FROM THE VIEWPOINT OF GRANULAR COMPUTING	1778
<i>Akinori Abe ; Yuki Hayashi</i>	
BIG-DATA-ENABLED MODELLING AND OPTIMIZATION OF GRANULAR SPEED-BASED VESSEL SCHEDULE RECOVERY PROBLEM	1786
<i>Fatemeh Cheraghchi ; Ibrahim Abualhaol ; Rafael Falcon ; Rami Abielmona ; Bijan Raahemi ; Emil Petriu</i>	

IMPROVING TEXT CLASSIFICATION WITH WORD EMBEDDING	1796
<i>Lihao Ge ; Teng-Sheng Moh</i>	
ON THE ROLE OF FEATURE SPACE GRANULATION IN FEATURE SELECTION PROCESSES	1806
<i>Marek Grzegorowski ; Andrzej Janusz ; Dominik Slezak ; Marcin Szczuka</i>	
QUASI-ERASABLE ITEMSET MINING	1816
<i>Tzung-Pei Hong ; Lu-Hung Chen ; Shyue-Liang Wang ; Chun-Wei Lin ; Bay Vo</i>	
SECURE INFORMATION FLOW AND FILE MOVEMENTS: A TOPOLOGICAL THEORY OF DISCRETIONARY ACCESS CONTROLS	1821
<i>Tsau Young T. Y. Lin ; Pierre Vachon</i>	
UNSUPERVISED DEEP EMBEDDING FOR NOVEL CLASS DETECTION OVER DATA STREAM	1830
<i>Ahmad M. Mustafa ; Gbadebo Ayoade ; Khaled Al-Naami ; Latifur Khan ; Kevin W. Hamlen ; Bhavani Thuraisingham ; Frederico Araujo</i>	
SCALABLE CYBER-SECURITY ANALYTICS WITH A NEW SUMMARY-BASED APPROXIMATE QUERY ENGINE	1840
<i>Dominik Slezak ; Agnieszka Chadzynska-Krasowska ; Joel Holland ; Piotr Synak ; Rick Glick ; Marcin Perkowski</i>	
MINING TEXT FOR DISEASE DIAGNOSIS IN HOSPITAL INFORMATION SYSTEM	1850
<i>Shusaku Tsumoto ; Tomohiro Kimura ; Haruko Iwata ; Shoji Hirano</i>	
NOISE SELF-FILTERING K-NEAREST NEIGHBORS ALGORITHMS	1860
<i>Shuyin Xia ; Guoyin Wang ; Yunsheng Liur ; Qun Liu ; Hong Yu</i>	
A PRELIMINARY STUDY ON DEEP LEARNING FOR PREDICTING SOCIAL INSURANCE PAYMENT BEHAVIOR	1866
<i>Josh Jia-Ching Ying ; Po-Yu Huang ; Chih-Kai Chang ; Don-Lin Yang</i>	
EFFECTS OF LANGUAGE PROCESSING IN TURKISH AUTHORSHIP ATTRIBUTION	1876
<i>Hayri Volkan Agun ; Sibel Yilmazel ; Ozgur Yilmazel</i>	
EVENT DETECTION FROM TIME-SERIES STREAMS USING DIRECTIONAL CHANGE AND DYNAMIC THRESHOLDS	1882
<i>Nora Alkhamees ; Maria Fasli</i>	
REAL-TIME LEXICON-BASED SENTIMENT ANALYSIS EXPERIMENTS ON TWITTER WITH A MILD (MORE INFORMATION, LESS DATA) APPROACH	1892
<i>Yusuf Arslan ; Aysenur Birturk ; Bekjan Djumabaev ; Dilek Küçük</i>	
A COMPARATIVE STUDY ON LEARNING TO RANK WITH COMPUTATIONAL METHODS	1898
<i>Inci Batmaz ; Pinar Karagoz ; Gulsah Serdar</i>	
SEMI-SUPERVISED LEARNING AND SOCIAL MEDIA TEXT ANALYSIS TOWARDS MULTI-LABELING CATEGORIZATION	1907
<i>Belainine Billal ; Alessandro Fonseca ; Fatiha Sadat ; Hakim Lounis</i>	
B3SAFIRBIYO: GENOMIC VARIANT ANALYSIS WITH BIG DATA TECHNOLOGIES	1917
<i>Tugçe Döngel ; Yasemin Timar</i>	
A DATA-DRIVEN APPROACH TO HELP UNDERSTANDING THE PREFERENCES OF PUBLIC TRANSPORT USERS	1926
<i>Vasco Furtado ; Elizabeth Furtado ; Carlos Caminha ; André Lopes ; Victor Dantas ; Caio Ponte ; Sofia Cavalcante</i>	
RECOVERING LOSS TO FOLLOWUP INFORMATION USING DENOISING AUTOENCODERS	1936
<i>Lovedeep Gondara ; Ke Wang</i>	
A RECOMMENDER MODEL BASED ON TRUST VALUE AND TIME DECAY: IMPROVE THE QUALITY OF PRODUCT RATING SCORE IN E-COMMERCE PLATFORMS	1946
<i>Muhittin Isik ; Hasan Dag</i>	
FOCUS LOCATION EXTRACTION FROM POLITICAL NEWS REPORTS WITH BIAS CORRECTION	1956
<i>Maryam Bahobj Imani ; Swarup Chandra ; Samuel Ma ; Latifur Khan ; Bhavani Thuraisingham</i>	
AUGMENTING WORD EMBEDDINGS THROUGH EXTERNAL KNOWLEDGE-BASE FOR BIOMEDICAL APPLICATION	1965
<i>Kishlay Jha ; Guangxu Xun ; Vishrawas Gopalakrishnan ; Aidong Zhang</i>	
BIG DATA IMPACT ON STABILITY AND RELIABILITY IMPROVEMENT OF SMART GRID	1975
<i>Shady S. Refaat ; Amira Mohamed ; Haitham Abu-Rub</i>	
A DEEP LEARNING MODEL FOR AIR QUALITY PREDICTION IN SMART CITIES	1983
<i>Ibrahim Kök ; Mehmet Ulvi Simsek ; Suat Özdemir</i>	
GRAPH-BASED INFORMATION EXPLORATION OVER STRUCTURED AND UNSTRUCTURED DATA	1991
<i>Giannis Koumoutsos ; Maria Fasli ; Ian Lewin ; David Milward</i>	

CONVOLUTIONAL NEURAL NETWORK FOR CLINICAL NARRATIVE CATEGORIZATION	2001
<i>Paula Lauren ; Guangzhi Qu ; Paul Watta</i>	
CLUSTOP: A CLUSTERING-BASED TOPIC MODELLING ALGORITHM FOR TWITTER USING WORD NETWORKS	2009
<i>Kwan Hui Lim ; Shanika Karunasekera ; Aaron Harwood</i>	
A NATURAL LANGUAGE NORMALIZATION APPROACH TO ENHANCE SOCIAL MEDIA TEXT REASONING	2019
<i>Long Hoang Nguyen ; Andrew Salopek ; Liang Zhao ; Fang Jin</i>	
USING META-LEARNING FOR MODEL TYPE SELECTION IN PREDICTIVE BIG DATA ANALYTICS	2027
<i>Mustafa V. Nural ; Hao Peng ; John A. Miller</i>	
WEATHER DATA ANALYSIS AND SENSOR FAULT DETECTION USING AN EXTENDED IOT FRAMEWORK WITH SEMANTICS, BIG DATA, AND MACHINE LEARNING	2037
<i>Aras Can Onal ; Omer Berat Sezer ; Murat Ozbayoglu ; Erdogan Dogdu</i>	
UNDERSTANDING WHAT AFFECTS CAREER PROGRESSION USING LINKEDIN AND TWITTER DATA	2047
<i>Yiming Pan ; Xuefeng Peng ; Tianran Hu ; Jiebo Luo</i>	
A DISTRIBUTED PROXIMAL GRADIENT DESCENT METHOD FOR TENSOR COMPLETION	2056
<i>T. Papastergiou ; V. Megalooikonomou</i>	
PREDICTING HIGH TAXI DEMAND REGIONS USING SOCIAL MEDIA CHECK-INS	2066
<i>Xuefeng Peng ; Yiming Pan ; Jiebo Luo</i>	
SLEEP-DEPRIVED FATIGUE PATTERN ANALYSIS USING LARGE-SCALE SELFIES FROM SOCIAL MEDIA	2076
<i>Xuefeng Peng ; Jiebo Luo ; Catherine Glenn ; Li-Kai Chi ; Jingyao Zhan</i>	
MATHEMATICAL PROGRAMMING FOR SOCIAL NETWORK ANALYSIS	2085
<i>Harun Pirim</i>	
UNSUPERVISED DEEP LEARNING FOR SUBSPACE CLUSTERING	2089
<i>Ali Sekmen ; Ahmet Bugra Koku ; Mustafa Parlaktuna ; Ayad Abdul-Malek ; Nagendrababu Vanamala</i>	
PRINCIPAL COORDINATE CLUSTERING	2095
<i>Ali Sekmen ; Akram Aldroubi ; Ahmet Bugra Koku ; Keaton Hamm</i>	
ESTIMATION OF PARAMETERS FOR THE FREE-FORM MACHINING WITH DEEP NEURAL NETWORK	2102
<i>Gokberk Serin ; M. Ugur Gudelek ; A. Murat Ozbayoglu ; Hakki Ozgur Unver</i>	
TOWARDS MAPREDUCE BASED BAYESIAN DEEP LEARNING NETWORK FOR MONITORING BIG DATA APPLICATIONS	2112
<i>M. Omair Shafiq ; Eric Torunski</i>	
MINED SEMANTIC ANALYSIS: A NEW CONCEPT SPACE MODEL FOR SEMANTIC REPRESENTATION OF TEXTUAL DATA	2122
<i>Walid Shalaby ; Wlodek Zadrozny</i>	
ONLINE VIDEO AD MEASUREMENT FOR POLITICAL SCIENCE RESEARCH	2132
<i>Adisak Sukul ; Baskar Gopalakrishnan ; Wallapak Tavanapong ; David A. M. Peterson</i>	
DXNAT — DEEP NEURAL NETWORKS FOR EXPLAINING NON-RECURRING TRAFFIC CONGESTION	2141
<i>Fangzhou Sun ; Abhishek Dubey ; Jules White</i>	
A FILTER-BASED FEATURE SELECTION MODEL FOR ANOMALY-BASED INTRUSION DETECTION SYSTEMS	2151
<i>Imtiaz Ullah ; Qusay H. Mahmoud</i>	
A HYBRID MODEL FOR ANOMALY-BASED INTRUSION DETECTION IN SCADA NETWORKS	2160
<i>Imtiaz Ullah ; Qusay H. Mahmoud</i>	
WHAT'S TRENDING TOMORROW, TODAY: USING EARLY ADOPTERS TO DISCOVER POPULAR POSTS ON TUMBLR	2168
<i>Daniel Xie ; Jiejun Xu ; Tsai-Ching Lu</i>	
HARVEY FLOODING RESCUE IN SOCIAL MEDIA	2177
<i>Zhou Yang ; Long Hoang Nguyen ; Joshua Stuve ; Guofeng Cao ; Fang Jin</i>	
A REVIEW ON CYBER SECURITY DATASETS FOR MACHINE LEARNING ALGORITHMS	2186
<i>Ozlem Yavanoglu ; Murat Aydos</i>	
ONE-SHOT LEARNING FOR FINE-GRAINED RELATION EXTRACTION VIA CONVOLUTIONAL SIAMESE NEURAL NETWORK	2194
<i>Jianbo Yuan ; Han Guo ; Zhiwei Jin ; Hongxia Jin ; Xianchao Zhang ; Jiebo Luo</i>	
SPEND PORTAL: LINKED DATA DISCOVERY USING SPARQL ENDPOINTS	2200
<i>Semih Yumusak ; Riza Emre Aras ; Elif Uysal ; Erdogan Dogdu ; Halife Kodaz ; Kasim Oztoprak</i>	

MODELING SELF-SERVICE MACHINE-LEARNING AGENTS FOR DISTRIBUTED STREAM PROCESSING	2203
<i>Philipp Zehnder ; Dominik Riemer</i>	
THE CYBERNETICS THOUGHT COLLECTIVE PROJECT: USING COMPUTATIONAL METHODS TO REVEAL INTELLECTUAL CONTEXT IN ARCHIVAL MATERIAL	2213
<i>Bethany G. Anderson ; Christopher J. Prom ; Kevin Hamilton ; James A. Hutchinson ; Mark Sammons ; Alex Dolski</i>	
IDENTIFYING EPOCHS IN TEXT ARCHIVES	2219
<i>Tobias Blanke ; Jon Wilson</i>	
GRAPHQL FOR ARCHIVAL METADATA: AN OVERVIEW OF THE EHRI GRAPHQL API	2225
<i>Mike Bryant</i>	
BUILDING NEW KNOWLEDGE FROM DISTRIBUTED SCIENTIFIC CORPUS: HERBADROP & EUROPEANA: TWO CONCRETE CASE STUDIES FOR EXPLORING BIG ARCHIVAL DATA	2231
<i>Pascal Dugenie ; Nuno Freire ; Daan Broeder</i>	
TOWARDS AUTOMATED QUALITY CURATION OF VIDEO COLLECTIONS FROM A REALISTIC PERSPECTIVE	2240
<i>Todd Goodall ; Maria Esteve ; Sandra Sweat ; Alan C. Bovik</i>	
WHAT CAN A KNOWLEDGE COMPLEXITY APPROACH REVEAL ABOUT BIG DATA AND ARCHIVAL PRACTICE?	2246
<i>Nicola Horsley</i>	
PROTECTING PRIVACY IN THE ARCHIVES: PRELIMINARY EXPLORATIONS OF TOPIC MODELING FOR BORN-DIGITAL COLLECTIONS	2251
<i>Tim Hutchinson</i>	
LINE DETECTION IN BINARY DOCUMENT SCANS: A CASE STUDY WITH THE INTERNATIONAL TRACING SERVICE ARCHIVES	2256
<i>Benjamin Charles Germain Lee</i>	
HEURISTICS FOR ASSESSING COMPUTATIONAL ARCHIVAL SCIENCE (CAS) RESEARCH: THE CASE OF THE HUMAN FACE OF BIG DATA PROJECT	2262
<i>Myeong Lee ; Yuheng Zhang ; Shiyun Chen ; Edel Spencer ; Jhon Dela Cruz ; Hyeonngi Hong ; Richard Marciano</i>	
A TYPOLOGY OF BLOCKCHAIN RECORDKEEPING SOLUTIONS AND SOME REFLECTIONS ON THEIR IMPLICATIONS FOR THE FUTURE OF ARCHIVAL PRESERVATION	2271
<i>Victoria L. Lemieux</i>	
AN INFRASTRUCTURE AND APPLICATION OF COMPUTATIONAL ARCHIVAL SCIENCE TO ENRICH AND INTEGRATE BIG DIGITAL ARCHIVAL DATA: USING TAIWAN INDIGENOUS PEOPLES OPEN RESEARCH DATA (TIPD) AS AN EXAMPLE	2279
<i>Ji-Ping Lin</i>	
AUTO-CATEGORIZATION METHODS FOR DIGITAL ARCHIVES	2288
<i>Nathaniel Payne ; Jason R. Baron</i>	
THE BLOCKCHAIN LITMUS TEST	2299
<i>T. D. Smith</i>	
COMPUTATIONAL CURATION OF A DIGITIZED RECORD SERIES OF WWII JAPANESE-AMERICAN INTERNMENT	2309
<i>William Underwood ; Richard Marciano ; Sandra Laib ; Carl Apgar ; Luis Beteta ; Waleed Falak ; Marisa Gilman ; Riss Hardcastle ; Keona Holden ; Yun Huang ; David Baasch ; Brittni Ballard ; Tricia Glaser ; Adam Gray ; Leigh Plummer ; Zeynep Diker ; Mayanka Jha ; Aakanksha Singh ; Namrata Walanj</i>	
TOWARDS A REQUIREMENTS ENGINEERING ARTEFACT MODEL IN THE CONTEXT OF BIG DATA SOFTWARE DEVELOPMENT PROJECTS: RESEARCH IN PROGRESS	2314
<i>Darlan Arruda ; Nazim H. Madhavji</i>	
PREDICTING OUTCOMES FOR BIG DATA PROJECTS: BIG DATA PROJECT DYNAMICS (BDPD): RESEARCH IN PROGRESS	2320
<i>David K. Becker</i>	
AGILE BIG DATA ANALYTICS: ANALYTICSOPS FOR DATA SCIENCE	2331
<i>Nancy W. Grady ; Jason A. Payne ; Huntley Parker</i>	
SAVING COSTS WITH A BIG DATA STRATEGY FRAMEWORK	2340
<i>Mike Lakoju ; Alan Serrano</i>	
DOES PAIR PROGRAMMING WORK IN A DATA SCIENCE CONTEXT? AN INITIAL CASE STUDY	2348
<i>Jeffrey S. Saltz ; Ivan Shamshurin</i>	
THE AMBIGUITY OF DATA SCIENCE TEAM ROLES AND THE NEED FOR A DATA SCIENCE WORKFORCE FRAMEWORK	2355
<i>Jeffrey S. Saltz ; Nancy W. Grady</i>	

MAKE ACCUMULATED DATA IN COMPANIES ELOQUENT BY SQL STATEMENT CONSTRUCTORS	2362
<i>Toshiyuki Shimon</i>	
ONLINE MINING FOR ASSOCIATION RULES AND COLLECTIVE ANOMALIES IN DATA STREAMS	2370
<i>Shaaban Abbady ; Cheng-Yuan Ke ; Jennifer Lavergne ; Jian Chen ; Vijay Raghavan ; Ryan Benton</i>	
ABC: A PRACTICABLE SKETCH FRAMEWORK FOR NON-UNIFORM MULTISSETS	2380
<i>Junzhi Gong ; Tong Yang ; Yang Zhou ; Dongsheng Yang ; Shigang Chen ; Bin Cui ; Xiaoming Li</i>	
HARNESSING THE POWER OF HASHTAGS IN TWEET ANALYTICS	2390
<i>Vibhuti Gupta ; Rattikorn Hewett</i>	
A STUDY OF A VIDEO ANALYSIS FRAMEWORK USING KAFKA AND SPARK STREAMING	2396
<i>Ayae Ichinose ; Atsuko Takefusa ; Hidemoto Nakada ; Masato Oguchi</i>	
TOWARDS A UNIFIED STORAGE AND INGESTION ARCHITECTURE FOR STREAM PROCESSING	2402
<i>Ovidiu-Cristian Marcu ; Alexandru Costan ; Gabriel Antoniu ; María S. Pérez-Hernández ; Radu Tudoran ; Stefano Bortoli ; Bogdan Nicolae</i>	
SMART DISTRIBUTED QUERY EXECUTION OVER DATA STREAMS	2408
<i>Salman Ahmed Shaikh ; Hiroyuki Kitagawa</i>	
RASP: REAL-TIME NETWORK ANALYTICS WITH DISTRIBUTED NOSQL STREAM PROCESSING	2414
<i>Georgios Touloupas ; Ioannis Konstantinou ; Nectarios Koziris</i>	
PREDICTING CONCEPT DRIFT VIA DYNAMIC NAÏVE BAYES	2420
<i>Qian Zhao ; Christian Klaue ; Chih Lai</i>	
LEVERAGING DISTRIBUTED BIG DATA STORAGE SUPPORT IN CLAAAS FOR WINGS WORKFLOW MANAGEMENT SYSTEM	2426
<i>Hadeel Alghamdi ; Farhana Zulkernine ; Patrick Martin</i>	
ONLINE MACHINE LEARNING FOR CLOUD RESOURCE PROVISIONING OF MICROSERVICE BACKEND SYSTEMS	2433
<i>Hanieh Alipour ; Yan Liu</i>	
TRILOGY: DATA PLACEMENT TO IMPROVE PERFORMANCE AND ROBUSTNESS OF CLOUD COMPUTING	2442
<i>Chin-Jung Hsu ; Vincent W. Freeh ; Flavio Villanustre</i>	
CLOSING THE LOOP — FINDING LUNG CANCER PATIENTS USING NLP	2452
<i>Bipin Karunakaran ; Debdipto Misra ; Kyle Marshall ; Dhruv Mathrawala ; Shravan Kethireddy</i>	
UNCOVERING THE EVOLUTION HISTORY OF DATA LAKES	2462
<i>Meike Klette ; Hannes Awolin ; Uta Störl ; Daniel Müller ; Stefanie Scherzinger</i>	
HIGHLY CONSOLIDATED SERVERS WITH CONTAINER-BASED VIRTUALIZATION	2472
<i>Joichiro Kon ; Naoki Mizusawa ; Ayaka Umezawa ; Saneyasu Yamaguchi ; Jian Tao</i>	
DYNAMIC DATA TRANSFORMATION FOR LOW LATENCY QUERYING IN BIG DATA SYSTEMS	2480
<i>Leandro Ordonez-Ante ; Thomas Vanhove ; Gregory Van Seghbroeck ; Tim Wauters ; Bruno Volckaert ; Filip De Turck</i>	
ICARUS: TOWARDS A MULTISTORE DATABASE SYSTEM	2490
<i>Marco Vogt ; Alexander Stiemer ; Heiko Schuldt</i>	
IMPROVING USER INTERACTION IN MOBILE-CLOUD DATABASE QUERY PROCESSING	2500
<i>Chenxiao Wang ; Jason Arenson ; Florian Helff ; Le Gruenwald ; Laurent D'Orazio</i>	
UNDERSTANDING AND IMPROVING DISK-BASED INTERMEDIATE DATA CACHING IN SPARK	2508
<i>Kaihui Zhang ; Yusuke Tanimura ; Hidemoto Nakada ; Hirotaka Ogawa</i>	
IMPROVING THE FUNCTIONALITY OF TAMURA DIRECTIONALITY ON SOLAR IMAGES	2518
<i>Azjin Ahmadzadeh ; Dustin J. Kempton ; Michael A. Schuh ; Rafal A. Angryk</i>	
PARALLEL COMPUTATION OF MAGNETIC FIELD PARAMETERS FROM HMI ACTIVE REGION PATCHES	2527
<i>Sunitha Basodi ; Berkay Aydın ; Rafal A. Angryk</i>	
ON THE PREDICTION OF >100 MEV SOLAR ENERGETIC PARTICLE EVENTS USING GOES SATELLITE DATA	2533
<i>Soukaina Filali Boubrahimi ; Berkay Aydın ; Petrus Martens ; Rafal Angryk</i>	
A TIME SERIES CLASSIFICATION-BASED APPROACH FOR SOLAR FLARE PREDICTION	2543
<i>Shah Muhammad Hamdi ; Dustin Kempton ; Ruizhe Ma ; Soukaina Filali Boubrahimi ; Rafal A. Angryk</i>	
MULTI-WAVELENGTH SOLAR EVENT DETECTION USING FASTER R-CNN	2552
<i>Ahmet Kucuk ; Berkay Aydın ; Rafal Angryk</i>	

IMPROVING EXPECTATION MAXIMIZATION ALGORITHM OVER STELLAR DATA	2559
<i>Hasan Kurban ; Can Kockan ; Mark Jenne ; Mehmet M. Dalkilic</i>	
SOLAR FLARE PREDICTION USING MULTIVARIATE TIME SERIES DECISION TREES	2569
<i>Ruizhe Ma ; Soukaina Filali Boubrahimi ; Shah Muhammad Hamdi ; Rafal A. Angryk</i>	
ACCELERATING SCIENTIFIC ALGORITHMS IN ARRAY DATABASES WITH GPUS	2579
<i>Simon Marcin ; André Csillaghy</i>	
IDENTIFYING AND MITIGATING RISKS TO THE QUALITY OF OPEN DATA IN THE POST-TRUTH ERA	2588
<i>Adrienne Colborne ; Michael Smit</i>	
GENERATIVE ADVERSARIAL NETWORKS FOR INCREASING THE VERACITY OF BIG DATA	2595
<i>Matthew L. Dering ; Conrad S. Tucker</i>	
AUGMENTATION AND EVALUATION OF TRAINING DATA FOR DEEP LEARNING	2603
<i>Junhua Ding ; Xinchuan Li ; Venkat N. Gudivada</i>	
IS DATA QUALITY ENOUGH FOR A CLINICAL DECISION?: APPLY MACHINE LEARNING AND AVOID BIAS	2612
<i>Kim Hee</i>	
DATA QUALITY CHALLENGES WITH MISSING VALUES AND MIXED TYPES IN JOINT SEQUENCE ANALYSIS	2620
<i>Alina Lazar ; Ling Jin ; C. Anna Spurlock ; Kesheng Wu ; Alex Sim</i>	
IMPROVING DATA QUALITY THROUGH HIGH PRECISION GENDER CATEGORIZATION	2628
<i>Daniel Müller ; Yiea-Funk Te ; Pratiksha Jain</i>	
COLLAPSING CORPORATE CONFUSION: LEVERAGING NETWORK STRUCTURES FOR EFFECTIVE ENTITY RESOLUTION IN RELATIONAL CORPORATE DATA	2637
<i>Tim Marple ; Bruce Desmarais ; Kevin L Young</i>	
TOWARD DATA QUALITY ANALYTICS IN SIGNATURE VERIFICATION USING A CONVOLUTIONAL NEURAL NETWORK	2644
<i>Shahab Tayeb ; Matin Pirouz ; Brittany Cozzens ; Richard Huang ; Maxwell Jay ; Kyle Khembunjong ; Sahar Paliskara ; Felix Zhan ; Mark Zhang ; Justin Zhan ; Shahram Latifi</i>	
AN IMPROVED P2P FILE SYSTEM SCHEME BASED ON IPFS AND BLOCKCHAIN	2652
<i>Yongle Chen ; Hui Li ; Kejiao Li ; Jiyang Zhang</i>	
THE ARCHITECTURE OF DISTRIBUTED STORAGE SYSTEM UNDER MIMIC DEFENSE THEORY	2658
<i>Hui Li ; Jiawei Hu ; Huaqun Ma ; Ting Huang</i>	
A SCHEDULING STRATEGY BASED ON MULTI-QUEUES OF CASSANDRA	2664
<i>Haopeng Li ; Hui Li</i>	
MDFS: A MIMIC DEFENSE THEORY BASED ARCHITECTURE FOR DISTRIBUTED FILE SYSTEM	2670
<i>Zhili Lin ; Kedan Li ; Hanxu Hou ; Xin Yang ; Hui Li</i>	
ON THE IMPLEMENTATION OF BRS CODES IN CEPH	2676
<i>Jiyang Zhang ; Hanxu Hou ; Kedan Li ; Hui Li</i>	
DETECTING POLARIZATION IN RATINGS: AN AUTOMATED PIPELINE AND A PRELIMINARY QUANTIFICATION ON SEVERAL BENCHMARK DATA SETS	2682
<i>Mahsa Badami ; Olfa Nasraoui ; Welong Sun ; Patrick Shafto</i>	
EVALUATING THE QUALITY OF GRAPH EMBEDDINGS VIA TOPOLOGICAL FEATURE RECONSTRUCTION	2691
<i>Stephen Bonner ; John Brennan ; Ibad Kureshi ; Georgios Theodoropoulos ; Andrew Stephen McGough ; Boguslaw Obara</i>	
USING SENTIMENT ANALYSIS TO EXPLORE THE DEGREE OF RISK IN SHARING ECONOMY	2701
<i>Wei-Lun Chang</i>	
PSEISMIC: A PERSONALIZED SELF-EXCITING POINT PROCESS MODEL FOR PREDICTING TWEET POPULARITY	2710
<i>Hsin-Yu Chen ; Cheng-Te Li</i>	
DETECTION OF PROFILE INJECTION ATTACKS IN SOCIAL RECOMMENDER SYSTEMS USING OUTLIER ANALYSIS	2714
<i>Anahita Davoudi ; Mainak Chatterjee</i>	
A BIG SOCIAL MEDIA DATA STUDY OF THE 2017 GERMAN FEDERAL ELECTION BASED ON SOCIAL SET ANALYSIS OF POLITICAL PARTY FACEBOOK PAGES WITH SOSEVI	2720
<i>Benjamin Flesch ; Ravi Vatrupu ; Raghava Rao Mukkamala</i>	
USING AN ASSET PRICE BUBBLE MODEL IN TWEET ANALYTICS	2730
<i>K. M. George</i>	

TOPIC LIFE CYCLE EXTRACTION FROM BIG TWITTER DATA BASED ON COMMUNITY DETECTION IN BIPARTITE NETWORKS.....	2740
<i>Takako Hashimoto ; Hiroshi Okamoto ; Tetsuji Kuboyama ; Kilho Shin</i>	
TICKET-PURCHASE BEHAVIOR UNDER THE EFFECTS OF MARKETING CAMPAIGNS ON FACEBOOK FAN PAGES.....	2746
<i>Hsiao-Wei Hu ; Ching-Han Cheng ; Yun-Chu Chung ; Chia-Yu Lee</i>	
LANGUAGE IDENTIFICATION IN MULTILINGUAL, SHORT AND NOISY TEXTS USING COMMON N-GRAMS.....	2752
<i>Dijana Kosmajac ; Vlado Keselj</i>	
CHARACTERIZATION OF DAILY TOURISM BEHAVIORS BASED ON PLACE SEQUENCE ANALYSIS FROM PHOTO SHARING WEBSITES.....	2760
<i>Thomas-Joseph Loiseau ; Sonia Djebali ; Thomas Raïmbault ; Bérengère Branchet ; Gaël Chareyron</i>	
DIGITAL CONTENT RECOMMENDATION SYSTEM USING IMPLICIT FEEDBACK DATA	2766
<i>Gang Wu ; Viswanathan Swaminathan ; Saayan Mitra ; Ratnesh Kumar</i>	
BIG SOCIAL DATA ANALYTICS FOR PUBLIC HEALTH: COMPARATIVE METHODS STUDY AND PERFORMANCE INDICATORS OF HEALTH CARE CONTENT ON FACEBOOK	2772
<i>Nadiya Straton ; Raghava Rao Mukkamala ; Ravi Vatrupu</i>	
OUTBOUND BEHAVIOR ANALYSIS THROUGH SOCIAL NETWORK DATA: A CASE STUDY OF CHINESE PEOPLE IN JAPAN.....	2778
<i>Tianqi Xia ; Xuan Song ; Dou Huang ; Satoshi Miyazawa ; Zipei Fan ; Renhe Jiang ; Ryosuke Shibasaki</i>	
TOWARDS ONLINE GRAPH PROCESSING WITH SPARK STREAMING	2787
<i>Tariq Abughofa ; Farhana Zulkernine</i>	
SUDS: SYSTEM FOR UNCERTAINTY DECISION SUPPORT.....	2795
<i>Maaïke De Boer ; Barry Nouwt ; Michael Van Bekkum</i>	
BIG DATA PROCESSING: IS THERE A FRAMEWORK SUITABLE FOR ECONOMISTS AND STATISTICIANS?	2804
<i>Giuseppe Bruno ; Demetrio Condello ; Alberto Falzone ; Andrea Luciani</i>	
A PERFORMANCE STUDY OF ASTERIXDB	2812
<i>Keren Ouaknine ; Michael Carey</i>	
PLUG AND PLAY BENCH: SIMPLIFYING BIG DATA BENCHMARKING USING CONTAINERS.....	2821
<i>Sheriffo Ceesay ; Adam Barker ; Blesson Varghese</i>	
ENHANCING THE MAPREDUCE TRAINING OF BP NEURAL NETWORKS BASED ON LOCAL WEIGHT MATRIX EVOLUTION	2829
<i>Wanghu Chen ; Xintian Li ; Jing Li ; Jianwu Wang</i>	
CLOUDEC: A MAPREDUCE-BASED ALGORITHM FOR CORRECTING ERRORS IN NEXT-GENERATION SEQUENCING BIG DATA.....	2836
<i>Wei-Chun Chung ; Jan-Ming Ho ; Chung-Yen Lin ; D. T. Lee</i>	
QUANTIFYING VOLUME, VELOCITY, AND VARIETY TO SUPPORT (BIG) DATA-INTENSIVE APPLICATION DEVELOPMENT.....	2843
<i>Rustem Dautov ; Salvatore Distefano</i>	
TULA: A DISK LATENCY AWARE BALANCING AND BLOCK PLACEMENT STRATEGY FOR HADOOP.....	2853
<i>Janakiram Dharanipragada ; Srikant Padala ; Balaji Kammili ; Vikram Kumar</i>	
EFFICIENT INCREMENTAL DATA ANALYTICS WITH APACHE SPARK.....	2859
<i>Sina Gholamian ; Wojciech Golab ; Paul A. S. Ward</i>	
A COMPARISON OF BIG DATA APPLICATION PROGRAMMING APPROACHES: A TRAVEL COMPANION CASE STUDY.....	2869
<i>Pei Guo ; Jianwu Wang ; Zhiyuan Chen</i>	
ADAPTIVE SCALABLE PIPELINES FOR POLITICAL EVENT DATA GENERATION	2879
<i>Andrew Halterman ; Jill Irvine ; Manar Landis ; Phanindra Jalla ; Yan Liang ; Christan Grant ; Mohiuddin Solaimani</i>	
IMBALANCE IN THE CLOUD: AN ANALYSIS ON ALIBABA CLUSTER TRACE	2884
<i>Chengzhi Lu ; Kejiang Ye ; Guoyao Xu ; Cheng-Zhong Xu ; Tongxin Bai</i>	
SCALING POINT SET REGISTRATION IN 3D ACROSS THREAD COUNTS ON MULTICORE AND HARDWARE ACCELERATOR PLATFORMS THROUGH AUTOTUNING FOR LARGE SCALE ANALYSIS OF SCIENTIFIC POINT CLOUDS	2893
<i>Piotr Luszczek ; Jakub Kurzak ; Ichitaro Yamazaki ; David Keffer ; Jack Dongarra</i>	
PERFORMANCE EVALUATION OF MULTIPLE SPORTS PLAYER TRACKING SYSTEM BASED ON GRAPH OPTIMIZATION	2903
<i>Yuri Nishikawa ; Hitoshi Sato ; Jun Ozawa</i>	

A PERFORMANCE STUDY OF BIG DATA ANALYTICS PLATFORMS.....	2911
<i>Pouria Pirzadeh ; Michael Carey ; Till Westmann</i>	
SCHEMA DESIGN SUPPORT FOR SEMI-STRUCTURED DATA: FINDING THE SWEET SPOT BETWEEN NF AND DE-NF	2921
<i>Vincent Reniers ; Dimitri Van Landuyt ; Ansar Rafique ; Wouter Joosen</i>	
A NOVEL COMPRESSION ALGORITHM DECISION METHOD FOR SPARK SHUFFLE PROCESS.....	2931
<i>Shanshan Huang ; Jungang Xu ; Renfeng Liu ; Husheng Liao</i>	
ECL-WATCH: A BIG DATA APPLICATION PERFORMANCE TUNING TOOL IN THE HPCC SYSTEMS PLATFORM.....	2941
<i>Lili Xu ; Edin Muharemagic ; Amy Apon</i>	
FINDING THE BEST BOX-COX TRANSFORMATION FROM MASSIVE DATASETS ON SPARK	2951
<i>Huayi Fang ; Baijian Yang ; Tonglin Zhang</i>	
COMMUNITY-BASED SELF GENERATION OF POLICIES AND PROCESSES FOR ASSETS: CONCEPTS AND RESEARCH DIRECTIONS	2961
<i>Elisa Bertino ; Geeth De Mel ; Alessandra Russo ; Seraphin Calo ; Dinesh Verma</i>	
RESEARCH CHALLENGES IN DYNAMIC POLICY-BASED AUTONOMOUS SECURITY	2970
<i>Seraphin Calo ; Emil Lupu ; Elisa Bertino ; Saritha Arunkumar ; Gregory Cirincione ; Brian Rivera ; Alan Cullen</i>	
MY (FAIR) BIG DATA.....	2974
<i>Tiziana Catarci ; Monica Scannapieco ; Marco Console ; Camil Demetrescu</i>	
LIGHTSPY: OPTICAL EAVESDROPPING ON DISPLAYS USING LIGHT SENSORS ON MOBILE DEVICES	2980
<i>Supriyo Chakraborty ; Wentao Ouyang ; Mani Srivastava</i>	
COMBINING SEMANTIC WEB AND IOT TO REASON WITH HEALTH AND SAFETY POLICIES.....	2990
<i>Emre Goynugur ; Murat Sensoy ; Geeth De Mel</i>	
IMPROVING DATA SHARING IN DATA RICH ENVIRONMENTS.....	2998
<i>Erisa Karafili ; Emil C. Lupu ; Alan Cullen ; Bill Williams ; Saritha Arunkumar ; Seraphin Calo</i>	
IDENTIFYING SENSOR ACCESSES FROM SERVICE DESCRIPTIONS.....	3006
<i>Antara Palit ; Mudhakar Srivatsa ; Raghu Ganti ; Christopher Simpkin</i>	
EDGE COMPUTING ARCHITECTURE FOR APPLYING AI TO IOT	3012
<i>Seraphin B. Calo ; Maroun Touna ; Dinesh C. Verma ; Alan Cullen</i>	
POLICY ENABLED CACHING FOR DISTRIBUTED AI	3017
<i>Dinesh C. Verma ; Graham Bent</i>	
CASE: BIG GEOSCIENCES DATA VALIDATION CHALLENGES AND ACHIEVEMENTS	3024
<i>Hussain Z. Al-Ajmi</i>	
WHY-DIFF: EXPLAINING DIFFERENCES AMONGST SIMILAR WORKFLOW RUNS BY EXPLOITING SCIENTIFIC METADATA.....	3031
<i>Priyaa Thavasimani ; Jacek Cala ; Paolo Missier</i>	
USING MACHINE LEARNING METHODS TO IDENTIFY ATROCITY PERPETRATORS	3042
<i>Benjamin E. Bagozzi ; Ore Koren</i>	
COMPARISON BETWEEN SPATIAL DISTRIBUTIONS OF TWEET BASE AND POPULATION IN JAPAN.....	3052
<i>Shouji Fujimoto ; Atushi Ishikawa ; Takayuki Mizuno</i>	
EVALUATING FUNDING PROGRAMS THROUGH NETWORK CENTRALITY MEASURES OF CO-AUTHOR NETWORKS OF TECHNICAL PAPERS	3058
<i>Masanori Fujita ; Hiroto Inoue ; Takao Terano</i>	
ANALYSIS OF TWITTER MESSAGES ABOUT THE OSAKA METROPOLIS PLAN IN JAPAN	3064
<i>Kouki Hayashi ; Eiichi Umehara ; Yuki Ogawa</i>	
ANALYZING REGIONAL CHARACTERISTICS OF LIVING ACTIVITIES OF ELDERLY PEOPLE FROM LARGE SURVEY DATA WITH PROBABILISTIC LATENT SPATIAL SEMANTIC STRUCTURE MODELING	3071
<i>Ayae Ide ; Kazuya Yamashita ; Yoichi Motomura ; Takao Terano</i>	
POSITION-SENSITIVE PROPAGATION OF INFORMATION ON SOCIAL MEDIA USING SOCIAL PHYSICS APPROACH	3078
<i>Akira Ishii ; Takayuki Mizuno ; Yasuko Kawahata</i>	
TIME DEPENDENT ANALYSIS OF FINANCIAL NETWORKS USING SUPERVISED LATENT FEATURE RELATIONAL MODELS	3086
<i>Shotaro Ito ; Koji Eguchi</i>	
A STATISTICAL ANALYSIS OF BEHAVIORAL BURSTS OCCURRING IN A SOCIAL NETWORKING GAME.....	3091
<i>Mitsuki Murase ; Masanori Takano ; Reiji Suzuki ; Takaya Arita</i>	

BIAS REDUCTION OF PEER INFLUENCE EFFECTS WITH LATENT COORDINATES AND COMMUNITY MEMBERSHIP	3098
<i>Daniel Rajchwald ; Natasha Markuzon ; Edoardo Airoldi</i>	
CROSS-NATIONAL MEASUREMENT OF POLARIZATION IN POLITICAL DISCOURSE: ANALYZING FLOOR DEBATE IN THE U.S. THE JAPANESE LEGISLATURES	3104
<i>Takuto Sakamoto ; Hiroki Takikawa</i>	
MINING SOCIAL MEDIA FOR DISASTER MANAGEMENT: LEVERAGING SOCIAL MEDIA DATA FOR COMMUNITY RECOVERY	3111
<i>Yuya Shibuya</i>	
WHEN DO USERS CHANGE THEIR PROFILE INFORMATION ON TWITTER?	3119
<i>Jinsei Shima ; Mitsuo Yoshida ; Kyoji Umemura</i>	
FACEBOOK AND PUBLIC HEALTH: A STUDY TO UNDERSTAND FACEBOOK POST PERFORMANCE WITH ORGANIZATIONS' STRATEGY	3123
<i>Nadiya Straton ; Ravi Vatraru ; Raghava Rao Mukkamala</i>	
DEVELOP METHOD TO PREDICT THE INCREASE IN THE NIKKEI VI INDEX	3133
<i>Hirohiko Suwa ; Yuki Ogawa ; Eiichi Umehara ; Kento Kakigi ; Keiichi Yasumoto ; Tatsuo Yamashita ; Kota Tsubouchi</i>	
ANALYSIS OF THE CHANGES IN LISTENING TRENDS OF A MUSIC STREAMING SERVICE	3139
<i>Masanori Takano ; Hiroki Mizukami ; Fujio Toriumi ; Makoto Takeuchi ; Kazuya Wada ; Masahiro Yasuda ; Ichiro Fukiida</i>	
POLITICAL POLARIZATION IN SOCIAL MEDIA: ANALYSIS OF THE "TWITTER POLITICAL FIELD" IN JAPAN	3143
<i>Hiroki Takikawa ; Kikuko Nagayoshi</i>	
ANALYSIS OF EXILE TRIBE IN THE MUSIC SCENE USING MATHEMATICAL MODEL OF HIT PHENOMENON	3151
<i>Toshimichi Wakabayashi ; Yasuko Kawahata ; Akira Ishii</i>	
RELATIONSHIPS BETWEEN MARKET IMPACT CHARACTERISTICS AND ORDER BOOK PROPERTIES	3156
<i>Kenta Yamada ; Takayuki Mizuno</i>	
DETECTING TWO TYPES OF SEASONAL WORDS USING SIMPLE AUTOCORRELATION ANALYSIS	3162
<i>Kenta Yamada</i>	
INFERENCE OF PERSONAL ATTRIBUTES FROM TWEETS USING MACHINE LEARNING	3168
<i>Take Yo ; Kazutoshi Sasahara</i>	
MANAGING MASSIVE MULTI-DIMENSIONAL ARRAY DATA WITH TILEDB: — INVITED DEMO PAPER	3175
<i>Jacob Bolewski ; Stavros Papadopoulos</i>	
GENERATING POLYSTORE INGESTION PLANS — A DEMONSTRATION WITH THE AWESOME SYSTEM	3177
<i>Subhasis Dasgupta ; Charles McKay ; Amarnath Gupta</i>	
POLYSTORE MATHEMATICS OF RELATIONAL ALGEBRA	3180
<i>Hayden Jananathan ; Ziqi Zhou ; Vijay Gadepally ; Dylan Hutchison ; Suna Kim ; Jeremy Kepner</i>	
QUERYING WEB POLYSTORES	3190
<i>Yasar Khan ; Antoine Zimmermann ; Alokkumar Jha ; Dietrich Rebbholz-Schuhmann ; Ratnesh Sahay</i>	
A NOVEL OBJECT PLACEMENT PROTOCOL FOR MINIMIZING THE AVERAGE RESPONSE TIME OF GET OPERATIONS IN DISTRIBUTED KEY-VALUE STORES	3196
<i>Antonios Makris ; Konstantinos Tserpes ; Dimosthenis Anagnostopoulos</i>	
SCIDB: AN ARRAY-NATIVE COMPUTATIONAL DATABASE FOR HETEROGENEOUS, MULTI-DIMENSIONAL DATA SETS	3206
<i>Jonathan Rivers</i>	
ENABLING QUERY PROCESSING ACROSS HETEROGENEOUS DATA MODELS: A SURVEY	3211
<i>Ran Tan ; Rada Chirkova ; Vijay Gadepally ; Timothy G. Mattson</i>	
AN APACHE CALCITE-BASED POLYSTORE VARIATION FOR FEDERATED QUERYING OF HETEROGENEOUS HEALTHCARE SOURCES	3221
<i>Ashwin Kumar Vajjantri ; Kunwar Deep Singh Toor ; Edmon Begoli ; Jack Bates</i>	
A DETECTION MECHANISM WITH TEXT MINING CROSS CORRELATION APPROACH	3228
<i>Jose Luis ; Guerrero Cusumano</i>	
TEXT MINING ANALYSIS OF WIND TURBINE ACCIDENTS: AN ONTOLOGY-BASED FRAMEWORK	3233
<i>Gurdal Ertek ; Xu Chi ; Allan N. Zhang ; Sobhan Asian</i>	

A MODEL FOR ANALYSING A DISRUPTED SUPPLY CHAIN'S TIME-TO-RECOVERY UNDER UNCERTAINTY	3242
<i>A. J. L. Lee ; D. Paul ; W. J. Yan ; A. N. Zhang ; Mark Goh</i>	
APPLICATION OF DEEP NEURAL NETWORK AND GENERATIVE ADVERSARIAL NETWORK TO INDUSTRIAL MAINTENANCE: A CASE STUDY OF INDUCTION MOTOR FAULT DETECTION	3248
<i>Yong Oh Lee ; Jun Jo ; Jongwoon Hwang</i>	
LEARNING AUTOMATA BASED METHOD FOR SOLVING DEMAND AND SUPPLY PROBLEM WITH PERIODIC BEHAVIORS	3254
<i>Haoye Lu ; Anand Srinivasan ; Amiya Nayak</i>	
FORECAST AND ANALYSIS OF FOOD DONATIONS USING SUPPORT VECTOR REGRESSION	3261
<i>Nigel Pugh ; Lauren B. Davis</i>	
ASSOCIATION ANALYSIS OF SUPPLY CHAIN RISK AND COMPANY SALES	3268
<i>Murat Mustafa Tunc ; Alexandru Valcov ; Allan N. Zhang ; Wenjing Yan ; Rong Wen</i>	
ADAPTIVE SPATIO-TEMPORAL MINING FOR ROUTE PLANNING AND TRAVEL TIME ESTIMATION	3278
<i>Rong Wen ; Wenjing Yan ; Allan N. Zhang</i>	
STREAMING ANALYTICS PROCESSING IN MANUFACTURING PERFORMANCE MONITORING AND PREDICTION	3285
<i>Yi-Hsin Wu ; Sheng-De Wang ; Li-Jung Chen ; Cheng-Juei Yu</i>	
PERFORMING LITERATURE REVIEW USING TEXT MINING, PART I: RETRIEVING TECHNOLOGY INFRASTRUCTURE USING GOOGLE SCHOLAR AND APIS	3290
<i>Dazhi Yang ; Allan N. Zhang ; Wenjing Yan</i>	
PERFORMING LITERATURE REVIEW USING TEXT MINING, PART II: EXPANDING DOMAIN KNOWLEDGE WITH ABBREVIATION IDENTIFICATION	3297
<i>Dazhi Yang ; Jihoon Hong</i>	
GPU-BASED PARALLEL ALGORITHM FOR GENERATING MASSIVE SCALE-FREE NETWORKS USING THE PREFERENTIAL ATTACHMENT MODEL	3302
<i>Maksudul Alam ; Kalyan S. Perumalla</i>	
A PARALLEL ALGORITHM FOR GENERATING A RANDOM GRAPH WITH A PRESCRIBED DEGREE SEQUENCE	3312
<i>Hasanuzzaman Bhuiyan ; Maleq Khan ; Madhav Marathe</i>	
DISCOVERING INTERESTING PATTERNS IN LARGE GRAPH CUBES	3322
<i>Florian Demesmaeker ; Amine Ghrab ; Siegfried Nijssen ; Sabri Skhiri</i>	
DISTRIBUTED MEMORY PARALLEL MARKOV RANDOM FIELDS USING GRAPH PARTITIONING	3332
<i>C. Heinemann ; T. Perciano ; D. Ushizima ; E. W. Bethel</i>	
A GENERALIZED INCREMENTAL BOTTOM-UP COMMUNITY DETECTION FRAMEWORK FOR HIGHLY DYNAMIC GRAPHS	3342
<i>Weiyi Liu ; Toyotaro Suzumura ; Lingli Chen ; Guangmin Hu</i>	
REGULAR DECOMPOSITION OF LARGE GRAPHS AND OTHER STRUCTURES: SCALABILITY AND ROBUSTNESS TOWARDS MISSING DATA	3352
<i>Hannu Reittu ; Ilkka Norros</i>	
STRIDER^R: MASSIVE AND DISTRIBUTED RDF GRAPH STREAM REASONING	3358
<i>Xiangnan Ren ; Olivier Curé ; Hubert Naacke ; Jérémy Lhez ; Ke Li</i>	
PRACTICAL APPROACH TO EVACUATION PLANNING VIA NETWORK FLOW AND DEEP LEARNING	3368
<i>Akira Tanaka ; Nozomi Hata ; Nariaki Tateiwa ; Katsuki Fujisawa</i>	
TECHNIQUES FOR EFFICIENT DETECTION OF RAPID WEATHER CHANGES AND ANALYSIS OF THEIR IMPACTS ON A HIGHWAY NETWORK	3378
<i>Adil Alim ; Aparna Joshi ; Feng Chen ; Catherine T. Lawson</i>	
SQL VERSUS NOSQL DATABASES FOR GEOSPATIAL APPLICATIONS	3388
<i>Elena Baralis ; Andrea Dalla Valle ; Paolo Garza ; Claudio Rossi ; Francesco Scullino</i>	
SPATIOTEMPORAL VISUALIZATION OF TRAFFIC PATHS USING COLOR SPACE TIME CURVE	3398
<i>Savitha Baskaran ; Shiaofoen Fang ; Shenhui Jiang</i>	
ALL IN ONE: ENCODING SPATIO-TEMPORAL BIG DATA IN XML, JSON, AND RDF WITHOUT INFORMATION LOSS	3406
<i>Peter Baumann ; Eric Hirschorn ; Joan Maso ; Vlad Merticariu ; Dimitar Misev</i>	
SPATEN: A SPATIO-TEMPORAL AND TEXTUAL BIG DATA GENERATOR	3416
<i>Thaleia Dimitra Doudali ; Ioannis Konstantinou ; Nectarios Koziris</i>	

MULTISCALE GRAPH THEORETICAL TOOLS REVEAL SUBTLE PATTERNS IN BIG GEOSPATIAL DATA	3422
<i>Ronald D. Hagan ; Charles A. Phillips ; Michael A. Langston ; Bradley J. Rhodes</i>	
OPTIMAL VIEWPOINT FINDING FOR 3D VISUALIZATION OF SPATIO-TEMPORAL VEHICLE TRAJECTORIES ON CAUTION CROSSROADS DETECTED FROM VEHICLE RECORDER BIG DATA	3426
<i>Masahiko Itoh ; Daisaku Yokoyama ; Masashi Toyoda ; Masaru Kitsuregawa</i>	
ROAD MAP EXTRACTION FROM SATELLITE IMAGERY USING CONNECTED COMPONENT ANALYSIS AND LANDSCAPE METRICS	3435
<i>Kulsawasd Jitkajornwanich ; Peerapon Vateekul ; Teerapong Panboonyuen ; Siam Lawawirojwong ; Siwapon Srisonphan</i>	
SCALABLE PARALLEL DATA LOADING IN SCIDB	3443
<i>Sangchul Kim ; Junhee Lee ; Taehoon Kim ; Bongki Moon</i>	
DISCOVERING DYNAMIC PATTERNS OF URBAN SPACE VIA SEMI-NONNEGATIVE MATRIX FACTORIZATION	3447
<i>Zhicheng Liu ; Jun Cao ; Junyan Yang ; Qiao Wang</i>	
IDENTIFYING COHERENT ANOMALIES IN MULTI-SCALE SPATIO-TEMPORAL DATA USING MARKOV RANDOM FIELDS	3454
<i>Adway Mitra</i>	
A TALE OF TWO CITIES: ANALYZING ROAD ACCIDENTS WITH BIG SPATIAL DATA	3461
<i>Rene Richard ; Suprio Ray</i>	
CHALLENGES AND TRENDS ABOUT SMART BIG GEOSPATIAL DATA: A POSITION PAPER	3471
<i>Victor Saquicela ; Luis M. Vilches-Blázquez ; Andrés Tello</i>	
TOWARDS DEVELOPMENT OF SPARK BASED AGRICULTURAL INFORMATION SYSTEM INCLUDING GEO-SPATIAL DATA	3476
<i>Purnima Shah ; Deepak Hiremath ; Sanjay Chaudhary</i>	
A MAP-BASED VISUAL ANALYSIS METHOD FOR PATTERNS DISCOVERY OF MOBILE LEARNING IN EDUCATION WITH BIG DATA	3482
<i>Dongbo Zhou ; Hao Li ; Sannyuya Liu ; Bo Song ; Tony Xiaohua Hu</i>	
BIG DATA MACHINE LEARNING USING APACHE SPARK MLLIB	3492
<i>Mehdi Assefi ; Ehsun Behravesht ; Guangchi Liu ; Ahmad P. Tafti</i>	
RETURN OF EXPERIENCE ON THE MEAN-SHIFT CLUSTERING FOR HETEROGENEOUS ARCHITECTURE USE CASE	3499
<i>Christophe Cérin ; Jean-Luc Gaudiot ; Mustapha Lebbah ; Fousté Yuehgoh</i>	
CLOUD BIG DATA DECISION SUPPORT SYSTEM FOR MACHINE LEARNING ON AWS: ANALYTICS OF ANALYTICS	3508
<i>Alex Kaplunovich ; Yelena Yesha</i>	
DIVIDE-AND-CONQUER STRATEGIES FOR LARGE-SCALE SIMULATIONS IN R	3517
<i>Hui Zhang ; Yiwen Zhong ; Juan Lin</i>	
MAP-SCAN NODE ACCELERATOR FOR BIG-DATA	3524
<i>Mihaela Malita ; Gheorghe M. Stefan</i>	
RANKED TIME SERIES MATCHING BY INTERLEAVING SIMILARITY DISTANCES	3530
<i>Cuong Nguyen ; Charles Lovering ; Rodica Neamtu</i>	
KERNEL BANDWIDTH SELECTION FOR SVDD: THE SAMPLING PEAK CRITERION METHOD FOR LARGE DATA	3540
<i>Sergiy Peredriy ; Deovrat Kakde ; Arin Chaudhuri</i>	
AN ONLINE SPATIO-TEMPORAL MODEL FOR INFERENCE AND PREDICTIONS OF TAXI DEMAND	3550
<i>Hong Yan ; Zhongqiang Zhang ; Jian Zou</i>	
MACHINE LEARNING FOR EARLY DETECTION OF AUTISM (AND OTHER CONDITIONS) USING A PARENTAL QUESTIONNAIRE AND HOME VIDEO SCREENING	3558
<i>Halim Abbas ; Ford Garberson ; Eric Glover ; Dennis P. Wall</i>	
ARTIFICIAL INTELLIGENCE APPLIED TO CHALLENGES IN THE FIELDS OF OPERATIONS AND CUSTOMER SUPPORT	3562
<i>Ravi Santosh Arvapally ; Hasan Hicsasmaz ; Wally Lo Faro</i>	
SEMANTIC SEARCH (INVITED TALK)	3570
<i>Ricardo Baeza-Yates</i>	
ARTIFICIAL INTELLIGENCE(AI), AUTOMATION, AND ITS IMPACT ON DATA SCIENCE	3571
<i>Richard Boire</i>	
A HYBRID BIPARTITE GRAPH BASED RECOMMENDATION ALGORITHM FOR MOBILE GAMES	3575
<i>Yong Cai ; Shaorong Liu ; Jinlong Hu ; Guihong Bai ; Shoubin Dong</i>	

ESTIMATING SKILL FUNGIBILITY AND FORECASTING SERVICES LABOR DEMAND	3583
<i>Brian Johnston ; Benjamin Zweig ; Michael Peran ; Charlie Wang ; Rachel Rosenfeld</i>	
INNOVATION IN BIG DATA ANALYTICS: APPLICATIONS OF MATHEMATICAL PROGRAMMING IN MEDICINE AND HEALTHCARE	3586
<i>Eva K. Lee</i>	
AUTOMATED KNOWLEDGE EXTRACTION FROM THE FEDERAL ACQUISITION REGULATIONS SYSTEM (FARS)	3596
<i>Srishty Saha ; Karuna P. Joshi ; Renee Frank ; Michael Aebig ; Jiayong Lin</i>	
A COMPARATIVE SEQUENCE ANALYSIS OF CAREER PATHS AMONG KNOWLEDGE WORKERS IN A MULTINATIONAL BANK	3604
<i>Paul Squires ; Harold G. Kaufman ; Julian Togelius ; Catalina M. Jaramillo</i>	
HITTING YOUR NUMBER OR NOT? A ROBUST & INTELLIGENT SALES FORECAST SYSTEM	3613
<i>Xin Xu Lei ; Tang Venkat Rangan</i>	
GOVERNANCE FRAMEWORK FOR ENTERPRISE ANALYTICS AND DATA	3623
<i>Atsushi Yamada ; Michael Peran</i>	
FORENSICS ANALYSIS OF WI-FI COMMUNICATION TRACES IN MOBILE DEVICES	3632
<i>Anja Evelyn Amundsen ; Kenneth M. Owens</i>	
IDENTIFYING EXTREMISM IN SOCIAL MEDIA WITH MULTI-VIEW CONTEXT-AWARE SUBSET OPTIMIZATION	3638
<i>Sreyasee Das Bhattacharjee ; Bala Venkatram Balantrapu ; William Tolone ; Ashit Talukder</i>	
EXTRACTING CYBER THREAT INTELLIGENCE FROM HACKER FORUMS: SUPPORT VECTOR MACHINES VERSUS CONVOLUTIONAL NEURAL NETWORKS	3648
<i>Isuf Deliu ; Carl Leichter ; Katrin Franke</i>	
EXPLORATORY STUDIES INTO FORENSIC LOGS FOR CRIMINAL INVESTIGATION USING CASE STUDIES IN INDUSTRIAL CONTROL SYSTEMS IN THE POWER SECTOR	3657
<i>Asif Iqbal ; Mathias Ekstedt ; Hanan Alobaidli</i>	
NEURAL REPUTATION MODELS LEARNED FROM PASSIVE DNS DATA	3662
<i>Pierre Lison ; Vasileios Mavroeidis</i>	
CYBER CRIME INVESTIGATIONS IN THE ERA OF BIG DATA	3672
<i>Andrii Shalaginov ; Jan William Johnsen ; Katrin Franke</i>	
TOPICAL BEHAVIOR PREDICTION FROM MASSIVE LOGS	3677
<i>Shih-Chieh Su</i>	
INTRODUCING DEEPBALANCE: RANDOM DEEP BELIEF NETWORK ENSEMBLES TO ADDRESS CLASS IMBALANCE	3684
<i>Peter Xenopoulos</i>	
A FIRST ESTIMATION OF THE PROPORTION OF CYBERCRIMINAL ENTITIES IN THE BITCOIN ECOSYSTEM USING SUPERVISED MACHINE LEARNING	3690
<i>Haohua Sun Yin ; Ravi Vatraru</i>	
FORENSIC DATABASE RECONSTRUCTION	3700
<i>Joshua Sablatura ; Bing Zhou</i>	
COUPLING EARLY WARNING SERVICES, CROWDSOURCING, AND MODELLING FOR IMPROVED DECISION SUPPORT AND WILDFIRE EMERGENCY MANAGEMENT	3705
<i>C. Bielski ; V. O'Brien ; C. Whitmore ; K. Ylinen ; I. Juga ; P. Nurmi ; J. Kilpinen ; I. Porras ; J. M. Sole ; P. Gamez ; M. Navarro ; A. Alikadic ; A. Gobbi ; C. Furlanello ; G. Zeug ; M. Weirathe ; J. Martinez ; R. Yuste ; S. Castro ; V. Moreno ; T. Velin ; C. Rossi</i>	
SUMMARIZATION OF EMERGENCY NEWS ARTICLES DRIVEN BY RELEVANCE FEEDBACK	3713
<i>Luca Cagliero</i>	
ALL IN A TWITTER: SELF-TUNING STRATEGIES FOR A DEEPER UNDERSTANDING OF A CRISIS TWEET COLLECTION	3722
<i>Evelina Di Corso ; Francesco Ventura ; Tania Cerquitelli</i>	
GAMIFIED CROWDSOURCING FOR DISASTER RISK MANAGEMENT	3727
<i>Antonella Frisiello ; Quynh Nhu Nguyen ; Claudio Rossi</i>	
A HEAT WAVE FORECAST SYSTEM FOR EUROPE	3734
<i>Andrea Gobbi ; Azra Alikadic ; Kaisa Ylinen ; Federico Angarano ; Cesare Furlanello</i>	
A LANGUAGE-AGNOSTIC APPROACH TO EXACT INFORMATIVE TWEETS DURING EMERGENCY SITUATIONS	3739
<i>Jacopo Longhini ; Claudio Rossi ; Claudio Casetti ; Federico Angarano</i>	
RIVER SEGMENTATION FOR FLOOD MONITORING	3746
<i>Laura Lopez-Fuentes ; Claudio Rossi ; Harald Skinnemoen</i>	

A COMPARISON OF CLASSIFICATION MODELS FOR NATURAL DISASTER AND CRITICAL EVENT DETECTION FROM NEWS	3750
<i>Tim Nugent ; Fabio Petroni ; Natraj Raman ; Lucas Carstens ; Jochen L. Leidner</i>	
OPTIMAL GEOSPATIAL VOLUNTEER ALLOCATION NEEDS REALISTIC DISTANCES	3760
<i>Jasmin Pielorz ; Matthias Prandtstetter ; Markus Straub ; Christoph H. Lampert</i>	
CROWD CONTROL AND EVACUATION GUIDANCE BASED ON SIMULATIONS	3764
<i>Tomoichi Takahashi ; Katsuki Ichinose</i>	
THE ROLE OF UNSTRUCTURED DATA IN REAL-TIME DISASTER-RELATED SOCIAL MEDIA MONITORING	3769
<i>Francesco Tarasconi ; Michela Farina ; Antonio Mazzei ; Alessio Bosca</i>	
ANALYZING SPATIAL DATA FROM TWITTER DURING A DISASTER	3779
<i>Luca Venturini ; Evelina Di Corso</i>	
COMPARISON OF DIFFERENT DRIVING STYLE ANALYSIS APPROACHES BASED ON TRIP SEGMENTATION OVER GPS INFORMATION	3784
<i>Marco Brambilla ; Paolo Mascetti ; Andrea Mauri</i>	
UNDERSTANDING DATA QUALITY: ENSURING DATA QUALITY BY DESIGN IN THE RAIL INDUSTRY	3792
<i>Qian Fu ; John M. Easton</i>	
TRACK GEOMETRY BIG DATA ANALYSIS: A MACHINE LEARNING APPROACH	3800
<i>Emmanuel Nii Martey ; Lasisi Ahmed ; Nii Attoh-Okine</i>	
APPLICATION OF MACHINE LEARNING FOR FUEL CONSUMPTION MODELLING OF TRUCKS	3810
<i>Federico Perrotta ; Tony Parry ; Luis C. Neves</i>	
PRIVACY-PRESERVING TRAJECTORY CLASSIFICATION OF DRIVING TRIP DATA BASED ON PATTERN DISCOVERY TECHNIQUES	3816
<i>Gene P. K. Wu ; Keith C. C. Chan</i>	
PREDICTIVE ANALYTICS FOR LITIGATION CASE MANAGEMENT	3826
<i>Jerzy Bala ; Michael Kellar ; Fred Ramberg</i>	
USING GOOGLE ANALYTICS TO SUPPORT CYBERSECURITY FORENSICS	3831
<i>Han Qin ; Kit Riehle ; Haozhen Zhao</i>	
A FEASIBILITY EXPERIMENT ON THE APPLICATION OF PREDICTIVE CODING TO INSTANT MESSAGING CORPORA	3835
<i>Thanasis Schoinas ; Ghulam Qadir</i>	
PATIENT-INDIVIDUAL MORPHOLOGICAL ANOMALY DETECTION IN MULTI-LEAD ELECTROCARDIOGRAPHY DATA STREAMS	3841
<i>Alexander Acker ; Florian Schmidt ; Anton Gulenko ; Reinhard Kietzmann ; Odej Kao</i>	
PREDICTING EFFICACY OF THERAPEUTIC SERVICES FOR AUTISM SPECTRUM DISORDER USING SCIENTIFIC WORKFLOWS	3847
<i>Fahima Bhuyan ; Shiyong Lu ; Ishiaq Ahmed ; Jia Zhang</i>	
A MULTIMEDIA BIG DATA RETRIEVAL FRAMEWORK TO DETECT DYSLEXIA AMONG CHILDREN	3857
<i>Elham Hassanain</i>	
MINING ACCOMPANYING RELATIONSHIPS BETWEEN DISEASES FROM PATIENT RECORDS	3861
<i>Wei Hong Lee ; En Tzu Wang ; Arbee L. P. Chen</i>	
EXPLAINABLE DATA-DRIVEN MODELING OF PATIENT SATISFACTION SURVEY DATA	3869
<i>Ning Liu ; Soundar Kumara ; Eric Reich</i>	
A MULTI-TASK MACHINE LEARNING APPROACH FOR COMORBID PATIENT PRIORITIZATION	3877
<i>Goutam Mylavarapu ; Johnson P Thomas</i>	
VISUALIZATION OF NON-METRIC RELATIONSHIPS BY ADAPTIVE LEARNING MULTIPLE MAPS T-SNE REGULARIZATION	3882
<i>Xianjun Shen ; Xianchao Zhu ; Xingpeng Jiang ; Li Gao ; Tingting He ; Xiaohua Hu</i>	
BIGNN: AN OPEN-SOURCE BIG DATA TOOLKIT FOCUSED ON BIOMEDICAL SENTENCE CLASSIFICATION	3888
<i>Ahmad P. Tafti ; Ehsun Behravesht ; Mehdi Assefi ; Eric Larose ; Jonathan Badger ; John Mayer ; Anhai Doan ; David Page ; Peggy Peissig</i>	
TOWARD PREDICTING MEDICAL CONDITIONS USING K-NEAREST NEIGHBORS	3897
<i>Shahab Tayeb ; Matin Pirouz ; Johann Sun ; Kaylee Hall ; Andrew Chang ; Jessica Li ; Connor Song ; Apoorva Chauhan ; Michael Ferra ; Theresa Sager ; Justin Zhan ; Shahram Latifi</i>	
A MEDICAL PRICE PREDICTION SYSTEM USING HIERARCHICAL DECISION TREES	3904
<i>Anuja Tike ; Sanket Tavarageri</i>	

HIGH DIMENSIONAL DATA PROCESSING FOR FETAL ACTIVITY EVALUATION	3914
<i>Julian Voicu ; Denis Kouamé</i>	
IVAR: INTERACTIVE VISUAL ANALYTICS OF RADIOMICS FEATURES FROM LARGE-SCALE MEDICAL IMAGES	3916
<i>Lina Yu ; Hengle Jiang ; Hongfeng Yu ; Chi Zhang ; Josiah McAllister ; Dandan Zheng</i>	
BIG DATA TECHNOLOGY AND ETHICS CONSIDERATIONS IN CUSTOMER BEHAVIOR AND CUSTOMER FEEDBACK MINING	3924
<i>Xin Deng</i>	
CUSTOMER CHURN PREDICTION IN AN INTERNET SERVICE PROVIDER	3928
<i>Duyen Do ; Phuc Huynh ; Phuong Vo ; Tu Vu</i>	
TRAINING ON THE POLES FOR REVIEW SENTIMENT POLARITY CLASSIFICATION	3934
<i>Michael Kranzlein ; Dan Chia-Tien Lo</i>	
UNDERSTANDING RATING BEHAVIOR BASED ON MORAL FOUNDATIONS: THE CASE OF YELP REVIEWS	3938
<i>Pegah Nokhiz ; Fengjun Li</i>	
A SCALABLE SEQUENTIAL PRINCIPAL COMPONENT ANALYSIS ALGORITHM (SEQPCA) WITH APPLICATION TO USER ACCESS CONTROL ANALYSIS	3946
<i>Yixuan Qiu ; Wutao Wei</i>	
TOWARDS AN ETHICAL APPLICATION OF CUSTOMER FEEDBACK DATA	3955
<i>Ross Smith</i>	
DYNAMIC BAYESIAN PREDICTIVE MODEL FOR BOX OFFICE FORECASTING	3958
<i>Wutao Wei ; Le Zhang ; Qi Ding ; Bingrou Zhou</i>	
A BIG DATA ANALYTICS FRAMEWORK FOR FORECASTING RARE CUSTOMER COMPLAINTS: A USE CASE OF PREDICTING MA MEMBERS' COMPLAINTS TO CMS	3965
<i>Donghui Wu</i>	
HETEROGENEOUS KNOWLEDGE TRANSFER VIA DOMAIN REGULARIZATION FOR IMPROVING CROSS-DOMAIN COLLABORATIVE FILTERING	3968
<i>Yizhou Zang ; Xiaohua Hu</i>	
IENVIRONMENT: A SOFTWARE PLATFORM FOR INTEGRATED ENVIRONMENTAL MONITORING AND MODELING OF SURFACE WATER	3975
<i>Paulo Alencar ; Donald Cowan ; Doug Mulholland ; Bruce Macvicar ; Simon Courtenay ; Stephen Murphy ; Fred McGarry</i>	
NEW DATA PARADIGMS: FROM THE CROWD AND BACK	3979
<i>Rumi Chunara</i>	
UNIFYING THE OPEN BIG DATA WORLD: THE POSSIBILITIES* OF APACHE BEAM	3981
<i>Holden Karau</i>	
DEEP LEARNING ENABLED NATIONAL CANCER SURVEILLANCE	3982
<i>Georgia Tourassi</i>	
PREPARING DATA MANAGERS TO SUPPORT OPEN OCEAN SCIENCE: REQUIRED COMPETENCIES, ASSESSED GAPS, AND THE ROLE OF EXPERIENTIAL LEARNING	3984
<i>Lee Wilson ; Adrienne Colborne ; Michael Smit</i>	
MODELING MULTIPLE SUBSKILLS BY EXTENDING KNOWLEDGE TRACING MODEL USING LOGISTIC REGRESSION	3994
<i>Xuan Zhou ; Wenjun Wu ; Yong Han</i>	
APPLICATION SPECIFIC TRAFFIC CONTROL USING NETWORK VIRTUALIZATION NODE IN LARGE-SCALE DISASTERS	4004
<i>Tsumugi Tairaku ; Akihiro Nakao ; Saneyasu Yamaguchi ; Masato Oguchi</i>	
AUTOMATIC DETECTION OF DNS MANIPULATIONS	4010
<i>Martino Trevisan ; Idilio Drago ; Marco Mellia ; Maurizio M. Munafò</i>	
MINING AND MODELING WEB TRAJECTORIES FROM PASSIVE TRACES	4016
<i>Luca Vassio ; Marco Mellia ; Flavio Figueiredo ; Ana Paula Couto Da Silva ; Jussara M. Almeida</i>	
AUTOMATIC TOPIC DISCOVERY OF ONLINE HOSPITAL REVIEWS USING AN IMPROVED LDA WITH VARIATIONAL GIBBS SAMPLING	4022
<i>Richard De Groof ; Haiping Xu</i>	
FRAGRANCE TO VECTOR AS SCENT TECHNOLOGY	4030
<i>Noriaki Koide ; Yu Ichifuji</i>	
CROSS-DATABASE MAMMOGRAPHIC IMAGE ANALYSIS THROUGH UNSUPERVISED DOMAIN ADAPTATION	4035
<i>Deepak Kumar ; Chetan Kumar ; Ming Shao</i>	
GUIDEME: ROUTES COORDINATION OF PARTICIPATING AGENTS IN MOBILE CROWD SENSING PLATFORMS	4043
<i>Christine Bassem ; Azer Bestavros</i>	

A WHOLE BUILDING FAULT DETECTION USING WEATHER BASED PATTERN MATCHING AND FEATURE BASED PCA METHOD	4050
<i>Yimin Chen ; Jin Wen</i>	
A MODEL FOR THE SOCIALLY SMART CITY PRACTICAL USES OF CITY-LEVEL SOCIO-ECONOMIC INDICATORS	4058
<i>Donald Cowan ; Paulo Alencar ; Kyle Young ; Bryan Smale ; Ryan Erb ; Fred McGarry</i>	
USING SOCIAL MEDIA PHOTOS TO IDENTIFY TOURISM PREFERENCES IN SMART TOURISM DESTINATION	4068
<i>Mickael Figueredo ; Nelio Cacho ; Antonio Thome ; Andrea Cacho ; Frederico Lopes ; Maria Araujo</i>	
SELF-ADAPTIVE AND RESILIENT URBAN NETWORKING INFRASTRUCTURE FOR DISASTERS AND SMART CITY SERVICES	4074
<i>Paul G. Flikkema ; Morgan Vigil-Hayes</i>	
DATA ANALYSIS ON TRAIN TRANSPORTATION DATA WITH NONNEGATIVE MATRIX FACTORIZATION	4080
<i>Kyoichi Ito ; Masaki Ito ; Kosuke Miyazaki ; Keishi Tanimoto ; Kaoru Sezaki</i>	
RELIABILITY ANALYSIS OF AN IOT-BASED SMART PARKING APPLICATION FOR SMART CITIES	4086
<i>Anderson Araújo ; Rubem Kalebe ; Gustavo Giraõ ; Itamir Filho ; Kayo Gonçalves ; Bianor Neto</i>	
ROAD MARKING BLUR DETECTION WITH DRIVE RECORDER	4092
<i>Makoto Kawano ; Kazuhiro Mikami ; Satoshi Yokoyama ; Takuro Yonezawa ; Jin Nakazawa</i>	
DATAFYING CITY: DETECTING AND ACCUMULATING SPATIO-TEMPORAL EVENTS BY VEHICLE-MOUNTED SENSORS	4098
<i>Yasue Kishino ; Koh Takeuchi ; Yoshinari Shirai ; Futoshi Naya ; Naonori Ueda</i>	
ANALYTICAL TOOLBOX FOR SMART CITY APPLICATIONS: GARBAGE COLLECTION LOG USE CASE	4105
<i>Takahiro Komamizu ; Jin Nakazawa ; Toshiyuki Amagasa ; Hiroyuki Kitagawa ; Hideyuki Tokuda</i>	
CITY EVENT DETECTION FROM SOCIAL MEDIA WITH NEURAL EMBEDDINGS AND TOPIC MODEL VISUALIZATION	4111
<i>Shuhua Liu ; Patrick Jansson</i>	
PROPOSING AN ACCESS GATE TO FACILITATE KNOWLEDGE EXCHANGE FOR SMART CITY SERVICES	4117
<i>Zohreh Pourzolfaghar ; Markus Helfert ; Viviana Angely Bastidas Melo ; Ahmad Khalilijafarabad</i>	
MM360: A GPS-ASSISTED 360-DEGREE VIDEO SHARING SYSTEM FOR PARTICIPATORY EVENTS	4123
<i>Naoya Shibahara ; Ryoma Kondo ; Masayuki Iwai</i>	
TOWARDS BUILDING A HYBRID MODEL FOR PREDICTING STOCK INDEXES	4128
<i>Jonathan Creighton ; Farhana H. Zulkernine</i>	
AGGLOMERATION, NETWORK AND URBAN DEVELOPMENT — — A STUDY ON NEWSPAPER CONNECTION NETWORK INDEX OF CITIES	4134
<i>Dongmei Guo ; Jialong Zheng ; Xiaolan Yang</i>	
AN AUGMENTED FAMA AND FRENCH THREE-FACTOR MODEL USING SOCIAL INTERACTION	4142
<i>Lin Huo ; Xiaoli Sun</i>	
STOCK PRICE FORECASTING USING SUPPORT VECTOR REGRESSION: BASED ON NETWORK BEHAVIOR DATA	4148
<i>Quan Jin ; Kun Guo ; Yi Sun</i>	
INSURANCE PREMIUM OPTIMIZATION USING MOTOR INSURANCE POLICIES — A BUSINESS GROWTH CLASSIFICATION APPROACH	4154
<i>Daniel Müller ; Yiea-Funk Te</i>	
PREDICTING BUSINESS PERFORMANCE THROUGH PATENT APPLICATIONS	4159
<i>Daniel Müller ; Yiea-Funk Te ; Pratiksha Jain</i>	
FORECASTING TOURIST ARRIVALS WITH MACHINE LEARNING AND INTERNET SEARCH INDEX	4165
<i>Shaolong Sun ; Shouyang Wang ; Yunjie Wei ; Xianduan Yang ; Kwok-Leung Tsui</i>	
A NEW TIME SERIES PREDICTION METHOD BASED ON COMPLEX NETWORK THEORY	4170
<i>Minggang Wang ; André L. M. Vilela ; Lixin Tian ; Hua Xu ; Ruijin Du</i>	
AN ENHANCED LGSA-SVM FOR S&P 500 INDEX FORECAST	4176
<i>Jinxin Wang ; Wei Shang ; Zhengyang Liu ; Shouyang Wang</i>	
CAN SEARCH DATA HELP FORECAST INFLATION? EVIDENCE FROM A 13-COUNTRY PANEL	4184
<i>Yunjie Wei ; Xun Zhang ; Shouyang Wang</i>	

INTEGRATING HETEROGENEOUS DATA SOURCES FOR TRAFFIC FLOW PREDICTION THROUGH EXTREME LEARNING MACHINE.....	4189
<i>Qingqing Zhang ; Darren Jian ; Rui Xu ; Wei Dai ; Ying Liu</i>	
THE CONSTRUCTION AND APPLICATION OF EXPECTATIONS INDEX ON MONETARY POLICY	4199
<i>Zheng Guihuan ; Yao Qikun ; Wang Xingfen ; Yang Zhou</i>	
BIG DATA PROCESSING: IS THERE A FRAMEWORK SUITABLE FOR ECONOMISTS AND STATISTICIANS?	4204
<i>Giuseppe Bruno ; Demetrio Condello ; Alberto Falzone ; Andrea Luciani</i>	
CLUSTER-OVERLAP ALGORITHM FOR ASSESSING PREPROCESSING CHOICES IN ENVIRONMENTAL SUSTAINABILITY	4212
<i>Anne M. Denton ; Arighna Roy</i>	
CRITICAL ENABLERS OF SUSTAINABLE WATER MANAGEMENT (SWM): TEXT EVIDENCES FROM 10 COUNTRIES	4221
<i>Chu-Hua Kuei ; Christian N. Madu ; Picheng Lee</i>	
CHARACTERIZATION OF CITIES BASED ON WORLD GRID SQUARE STATISTICS ABOUT SPECIFIC PROPERTIES	4228
<i>Aki-Hiro Sato</i>	
WORLD GRID SQUARE CODES: DEFINITION AND AN EXAMPLE OF WORLD GRID SQUARE DATA	4238
<i>Aki-Hiro Sato ; Shoki Nishimura ; Hiroe Tsubaki</i>	
STATISTICAL ANALYSIS OF HOTEL PLAN POPULARITY IN REGIONAL TOURIST AREAS	4248
<i>Hiroshi Tsuda ; Masakazu Ando ; Yu Ichifuji</i>	
SUSTAINABLE BLOCKCHAIN-ENABLED SERVICES: SMART CONTRACTS	4255
<i>Craig Wright ; Antoaneta Serguieva</i>	
DEVELOPING SUSTAINABLE TRADING STRATEGIES USING DIRECTIONAL CHANGES WITH HIGH FREQUENCY DATA	4265
<i>Ailun Ye ; V. L. Raju Chinthalapati ; Antoaneta Serguieva ; Edward Tsang</i>	
SARGS METHOD FOR DISTRIBUTED ACTIONABLE PATTERN MINING USING SPARK	4272
<i>Arunkumar Bagavathi ; Pranava Mummoju ; Katarzyna Tarnowska ; Angelina A. Tzacheva ; Zbigniew W. Ras</i>	
VEHICLE PATH ESTIMATION USING DUAL-LEVEL CLUSTERING AND MULTI-SOURCE PREDICTION.....	4282
<i>I-Cheng Chang ; Yudi Pratama Halim ; Chun-Man Lin</i>	
COMBINING PATTERN MATCHING WITH WORD EMBEDDINGS FOR THE EXTRACTION OF EXPERIMENTAL VARIABLES FROM SCIENTIFIC LITERATURE	4287
<i>Helena F. Deus ; Corey Harper ; Darin McBeath ; Ron Daniel</i>	
OCEAN SURFACE CURRENT PREDICTION BASED ON HF RADAR OBSERVATIONS USING TRAJECTORY-ORIENTED ASSOCIATION RULE MINING	4293
<i>Kulsawasd Jitkajornwanich ; Peerapon Vateekul ; Upa Gupta ; Teeranai Kormongkolkul ; Arnon Jirakittayakorn ; Siam Lawawirojwong ; Siwapon Srisonphan</i>	
A DISTRIBUTED PIPELINE FOR DIDSON DATA PROCESSING.....	4301
<i>Liling Li ; Tyler Danner ; Jesse Eickholt ; Erin McCann ; Kevin Pangle ; Nicholas Johnson</i>	
DEEP MODEL STYLE: CROSS-CLASS STYLE COMPATIBILITY FOR 3D FURNITURE WITHIN A SCENE	4307
<i>Tse-Yu Pan ; Yi-Zhu Dai ; Wan-Lun Tsai ; Min-Chun Hu</i>	
IMPROVING ARABIC SENTIMENT ANALYSIS WITH SENTIMENT-SPECIFIC EMBEDDINGS	4314
<i>A. Aziz Altowayan ; Ashraf Elnagar</i>	
DIFFERENCES IN EMOJI SENTIMENT PERCEPTION BETWEEN READERS AND WRITERS	4321
<i>Jose Berengueres ; Dani Castro</i>	
TOPIC MODELLING ENRICHED LSTM MODELS FOR THE DETECTION OF NOVEL AND EMERGING NAMED ENTITIES FROM SOCIAL MEDIA	4329
<i>Patrick Jansson ; Shuhua Liu</i>	
AN ENTITY DISAMBIGUATION METHOD BASED ON LEADERRANK.....	4337
<i>Bingjing Jia ; Bin Wu ; Jinna Lv ; Pengpeng Zhou ; Yao Bu ; Ying Xing</i>	
IDENTIFYING EMERGENCY STAGES IN FACEBOOK POSTS OF POLICE DEPARTMENTS WITH CONVOLUTIONAL AND RECURRENT NEURAL NETWORKS AND SUPPORT VECTOR MACHINES	4343
<i>Nicolai Pogrebnyakov ; Edgar Maldonado</i>	
#ANOREXIA, #ANAREXIA, #ANAREXYIA: CHARACTERIZING ONLINE COMMUNITY PRACTICES WITH ORTHOGRAPHIC VARIATION.....	4353
<i>Ian Stewart ; Stevie Chancellor ; Munmun De Choudhury ; Jacob Eisenstein</i>	

CROSSING THE STREAMS: FUZZ TESTING WITH USER INPUT	4362
<i>Joseph A. Cottam ; Leslie Blaha ; Dimitri Zarzhitsky ; Mathew Thomas ; Elliott Skonski</i>	
IMPROVING CLASSIFICATION ACCURACY IN CROWDSOURCING THROUGH HIERARCHICAL REORGANIZATION	4372
<i>Xiaoni Duan ; Keishi Tajima</i>	
CROWD-BASED BEST-EFFORT NUMBER ESTIMATION	4375
<i>Yuzuki Furuhashi ; Masaki Matsubara ; Atsuyuki Morishima</i>	
[RESEARCH PAPER] FORMALIZING INTERRUPTIBLE ALGORITHMS FOR HUMAN OVER-THE-LOOP ANALYTICS	4378
<i>Austin Graham ; Yan Liang ; Le Gruenwald ; Christan Grant</i>	
CLARIFYING THE TRANSITION OF WORKLOAD FOR VICTIMS LIFE RECONSTRUCTION SUPPORT PROGRAMS IN AFFECTED LOCAL GOVERNMENTS USING THE VICTIMS MASTER DATABASE — COMPARISON BETWEEN THE 2007 CHUETSU-OKI EARTHQUAKE AND THE 2016 KUMAMOTO EARTHQUAKE-	4384
<i>Munenari Inoguchi ; Keiko Tamura ; Kei Horie ; Haruo Hayashi</i>	
ACTIVE PREFERENCE LEARNING FOR GENERATIVE ADVERSARIAL NETWORKS	4389
<i>Masahiro Kazama ; Viviane Takahashi</i>	
A CROWD-IN-THE-LOOP APPROACH FOR GENERATING CONFERENCE PROGRAMS WITH MICROTASKS	4394
<i>Naoki Kobayashi ; Masaki Matsubara ; Keishi Tajima ; Atsuyuki Morishima</i>	
METHOD TO GENERATE DISASTER-DAMAGE MAP USING 3D PHOTOMETRY AND CROWD SOURCING	4397
<i>Koyo Kobayashi ; Hidehiko Shishido ; Yoshinari Kameda ; Itaru Kitahara</i>	
IMPLICIT ORDER JOIN: JOINING LOG DATA WITH PROPERTY DATA BY DISCOVERING IMPLICIT ORDER-ORIENTED KEYS WITH HUMAN ASSISTANCE	4400
<i>Takahiro Komamizu ; Toshiyuki Amagasa ; Hiroyuki Kitagawa</i>	
CONCEPTUAL DESIGN FOR COMPREHENSIVE RESEARCH SUPPORT PLATFORM: SUCCESSFUL RESEARCH DATA MANAGEMENT GENERATING BIG DATA FROM LITTLE DATA	4407
<i>Mamiko Matsubayashi ; Keiko Kurata</i>	
A TRADE-OFF BETWEEN ESTIMATION ACCURACY OF WORKER QUALITY AND TASK COMPLEXITY	4410
<i>Yoshitaka Matsuda ; Yu Suzuki ; Satoshi Nakamura</i>	
COLLABORATIVE FILTERING AND RATING AGGREGATION BASED ON MULTICRITERIA RATING	4417
<i>Hiroki Morise ; Satoshi Oyama ; Masahito Kurihara</i>	
TOWARDS PREDICTING TASK PERFORMANCE FROM EEG SIGNALS	4423
<i>Michalis Papakostas ; Konstantinos Tsiakas ; Theodoros Giannakopoulos ; Fillia Makedon</i>	
PROACTIVE PRESERVATION OF WORLD HERITAGE BY CROWDSOURCING AND 3D RECONSTRUCTION TECHNOLOGY	4426
<i>Hidehiko Shishido ; Yutaka Ito ; Youhei Kawamura ; Toshiya Matsui ; Atsuyuki Morishima ; Itaru Kitahara</i>	
USING CATEGORIZED WEB BROWSING HISTORY TO ESTIMATE THE USER'S LATENT INTERESTS FOR WEB ADVERTISEMENT RECOMMENDATION	4429
<i>Panote Siriaraya ; Yuriko Yamaguchi ; Mimpei Morishita ; Yoichi Inagaki ; Reyn Nakamoto ; Jianwei Zhang ; Junichi Aoi ; Shinsuke Nakajima</i>	
“DEKATSU” ACTIVITY OF DATA AND SERVICE COLLABORATION AMONG PRIVATE COMPANIES AND ACADEMIC INSTITUTIONS FOR TOKYO METROPOLITAN RESILIENCE PROJECT	4435
<i>Keiko Tamura ; Naoshi Hirata</i>	
LINK BEFORE YOU SHARE: MANAGING PRIVACY POLICIES THROUGH BLOCKCHAIN	4438
<i>Agniva Banerjee ; Karuna Pande Joshi</i>	
AUTOMATED MICROSOFT OFFICE MACRO MALWARE DETECTION USING MACHINE LEARNING	4448
<i>Ruth Bearden ; Dan Chai-Tien Lo</i>	
FIGHTING FAKE NEWS SPREAD IN ONLINE SOCIAL NETWORKS: ACTUAL TRENDS AND FUTURE RESEARCH DIRECTIONS	4453
<i>Alina Campan ; Alfredo Cuzzocrea ; Traian Marius Truta</i>	
IMPACT OF SECURITY AWARENESS TRAINING ON PHISHING CLICK-THROUGH RATES	4458
<i>Anthony Carella ; Murat Kotsoev ; Traian Marius Truta</i>	
DATA MASKING TECHNIQUES FOR NOSQL DATABASE SECURITY: A SYSTEMATIC REVIEW	4467
<i>Alfredo Cuzzocrea ; Hossain Shahriar</i>	

TOR TRAFFIC ANALYSIS AND DETECTION VIA MACHINE LEARNING TECHNIQUES	4474
<i>Alfredo Cuzzocrea ; Fabio Martinelli ; Francesco Mercaldo ; Gianni Vercelli</i>	
MODELING USER COMMUNITIES FOR IDENTIFYING SECURITY RISKS IN AN ORGANIZATION	4481
<i>Anirban Das ; Min-Yi Shen ; Jisheng Wang</i>	
EFFICIENT AND PRIVATE APPROXIMATIONS OF DISTRIBUTED DATABASES CALCULATIONS	4487
<i>Philip Derbeko ; Shlomi Dolev ; Ehud Gudes ; Jeffrey D. Ullman</i>	
COLLABORATIVE CACHING TECHNIQUES FOR PRIVACY-PRESERVING LOCATION-BASED SERVICES IN PEER-TO-PEER ENVIRONMENTS.....	4497
<i>Kangsoo Jung ; Seog Park</i>	
SECURE POWER SCHEDULING AUCTION FOR SMART GRIDS USING HOMOMORPHIC ENCRYPTION	4507
<i>Haya Shajaiah ; Ahmed Abdelhadi ; Charles Clancy</i>	
A TOP-DOWN K-ANONYMIZATION IMPLEMENTATION FOR APACHE SPARK	4513
<i>Ugur Sopaoglu ; Osman Abul</i>	
SECURING THE POSITIONING SIGNALS OF AUTONOMOUS VEHICLES.....	4522
<i>Shahab Tayeb ; Matin Pirouz ; Gabriel Esguerra ; Kimiya Ghobadi ; Jimson Huang ; Robin Hill ; Derwin Lawson ; Stone Li ; Tiffany Zhan ; Justin Zhan ; Shahram Latifi</i>	
USER-PROFILE-BASED ANALYTICS FOR DETECTING CLOUD SECURITY BREACHES	4529
<i>Trishita Tiwari ; Ata Turk ; Alina Oprea ; Katzalin Olcoz ; Ayse K. Coskun</i>	
EVENT CLUSTERING & EVENT SERIES CHARACTERIZATION ON EXPECTED FREQUENCY	4536
<i>Conrad M. Albrecht ; Marcus Freitag ; Theodore G. Van Kessel ; Siyuan Lu ; Hendrik F. Hamann</i>	
PETROLEUM ANALYTICS LEARNING MACHINE' FOR OPTIMIZING THE INTERNET OF THINGS OF TODAY'S DIGITAL OIL FIELD-TO-REFINERY PETROLEUM SYSTEM	4542
<i>Roger N. Anderson</i>	
DEVELOPING AN EDGE COMPUTING PLATFORM FOR REAL-TIME DESCRIPTIVE ANALYTICS.....	4546
<i>Hung Cao ; Monica Wachowicz ; Sangwhan Cha</i>	
ENERGY EFFICIENCY DRIVEN BY A STORAGE MODEL AND ANALYTICS ON A MULTI-SYSTEM SEMANTIC INTEGRATION.....	4555
<i>Domitille Couloumb ; Charbel El Kaed ; Ayush Garg ; Chris Healey ; Jonathan Healey ; Stuart Sheehan</i>	
DATA DRIVEN MODELING FOR ENERGY CONSUMPTION PREDICTION IN SMART BUILDINGS.....	4562
<i>Aurora González-Vidal ; Alfonso P. Ramallo-González ; Fernando Terroso-Sáenz ; Antonio Skarmeta</i>	
MACHINE LEARNING AND AIR QUALITY MODELING.....	4570
<i>Christoph A. Keller ; Mathew J. Evans ; J. Nathan Kutz ; Steven Pawson</i>	
A LOW MAINTENANCE PARTICLE POLLUTION SENSING SYSTEM USING THE MINIMUM AIRFLOW PARTICLE COUNTER (MAPC).....	4577
<i>Ted Van Kessel ; Ramachandran Muralidhar ; Josephine B. Chang ; Jun-Song Wang ; Michael Schappert ; Hendrik F. Hamann</i>	
DISTRIBUTED WIRELESS SENSING FOR FUGITIVE METHANE LEAK DETECTION.....	4583
<i>Levente J. Klein ; Theodore Van Kessel ; Dhruv Nair ; Ramachandran Muralidhar ; Nigel Hinds ; Hendrik Hamann ; Norma Sosa</i>	
USING BIG DATA ANALYTICS AND IOT PRINCIPLES TO KEEP AN EYE ON UNDERGROUND INFRASTRUCTURE	4592
<i>Joshua Lieberman ; Alan Leidner ; George Percivall ; Carsten Rönsdorf</i>	
UNDERSTANDING THE IMPACT OF LOSSY COMPRESSIONS ON IOT SMART FARM ANALYTICS.....	4602
<i>Aekyeung Moon ; Jaeyoung Kim ; Jialing Zhang ; Hang Liu ; Seung Woo Son</i>	
MEASURES OF NETWORK CENTRICITY FOR EDGE DEPLOYMENT OF IOT APPLICATIONS	4612
<i>Dinesh C. Verma ; Geeth De Mel</i>	
SOURCE CHARACTERIZATION OF AIRBORNE EMISSIONS USING A SENSOR NETWORK: EXAMINING THE IMPACT OF SENSOR QUALITY, QUANTITY, AND WIND CLIMATOLOGY	4621
<i>Xiaochi Zhou ; Vinicius Amaral ; John D. Albertson</i>	
SENTIMENT ANALYSIS VIA MULTI-LAYER PERCEPTRON TRAINED BY META-HEURISTIC OPTIMISATION.....	4630
<i>Dabiah Ahmed Alboaneen ; Huaglory Tianfield ; Yan Zhang</i>	

DETECTION OF HACKING BEHAVIORS AND COMMUNICATION PATTERNS ON SOCIAL MEDIA	4636
<i>Olga Babko-Malaya ; Rebecca Cathey ; Steve Hinton ; David Maimon ; Taissa Gladkova</i>	
IMPROVING CYBER-ATTACK PREDICTIONS THROUGH INFORMATION FORAGING	4642
<i>Adam Dalton ; Bonnie Dorr ; Leon Liang ; Kristy Hollingshead</i>	
TWITTER-ENHANCED ANDROID MALWARE DETECTION	4648
<i>Jordan Deloach ; Doina Caragea</i>	
DERIVING CYBER USE CASES FROM GRAPH PROJECTIONS OF CYBER DATA REPRESENTED AS BIPARTITE GRAPHS	4658
<i>Mohammed Eslami ; George Zheng ; Hamed Eramian ; Georgiy Levchuk</i>	
BINARY MALWARE IMAGE CLASSIFICATION USING MACHINE LEARNING WITH LOCAL BINARY PATTERN	4664
<i>Jhu-Sin Luo ; Dan Chia-Tien Lo</i>	
ON THE RELEVANCE OF SOCIAL MEDIA PLATFORMS IN PREDICTING THE VOLUME AND PATTERNS OF WEB DEFAACEMENT ATTACKS	4668
<i>David Maimon ; Andrew Fukuda ; Steve Hinton ; Olga Babko-Malaya ; Rebecca Cathey</i>	
TOWARDS A DEFINITION OF CYBERSPACE TACTICS, TECHNIQUES AND PROCEDURES	4674
<i>Fernando Maymí ; Robert Bixler ; Randolph Jones ; Scott Lathrop</i>	
DNS GRAPH MINING FOR MALICIOUS DOMAIN DETECTION	4680
<i>Hau Tran ; An Nguyen ; Phuong Vo ; Tu Vu</i>	
NETWORK INTRUSION DETECTION USING WORD EMBEDDINGS	4686
<i>Xiaoyan Zhuo ; Jialing Zhang ; Seung Woo Son</i>	
BUILDING INDUSTRY NETWORK BASED ON BUSINESS TEXT: CORPORATE DISCLOSURES AND NEWS	4696
<i>Sung Whan Jeon ; Hye Jin Lee ; Sungzoon Cho</i>	
PREDICTING STOCK MOVEMENT DIRECTION WITH MACHINE LEARNING: AN EXTENSIVE STUDY ON S&P 500 STOCKS	4705
<i>Yang Jiao ; Jérémie Jakubowicz</i>	
CREDIT DECISION TOOL USING MOBILE APPLICATION DATA FOR MICROFINANCE IN AGRICULTURE	4714
<i>Naomi Simumba ; Suguru Okami ; Naohiko Kohtake</i>	
ANALYSIS OF NATIONAL ELECTION USING MATHEMATICAL MODEL OF HIT PHENOMENON	4722
<i>Masanori Ajito ; Yasuko Kawahata ; Akira Ishii</i>	
TOWARDS A BIG DATA REQUIREMENTS ENGINEERING ARTEFACT MODEL IN THE CONTEXT OF BIG DATA SOFTWARE DEVELOPMENT PROJECTS: POSTER EXTENDED ABSTRACT	4725
<i>Darlan Arruda ; Nazim H. Madhavji</i>	
BIG DATA ANALYSIS OF YOUTH TOBACCO SMOKING TRENDS IN THE UNITED STATES	4727
<i>Shilpa Balan ; Nishant Shristiraj ; Vrunda Shah ; Anusha Manjappa</i>	
TOWARDS SCALABLE KERNEL MACHINES FOR STREAMING DATA ANALYTICS	4730
<i>Shaunak D. Bopardikar ; George S. Eskander Ekladious</i>	
LARGE SCALE APP RECOMMENDATION IN ANT FINANCIAL	4733
<i>Chaochao Chen ; Xinxing Yang ; Li Wang ; Jun Zhou ; Xiaolong Li</i>	
SOCIAL MEDIA BASED NPL SYSTEM TO FIND AND RETRIEVE ARM DATA: CONCEPT PAPER	4736
<i>Ranjeet Devarakonda ; Michael Giansiracusa ; Jitendra Kumar ; Harold Shanafield</i>	
TOWARDS A DISTRIBUTED INFRASTRUCTURE FOR DATA-DRIVEN DISCOVERIES & ANALYSIS	4738
<i>Mohammed Elshambakey ; Mohamed Khalefa ; William J. Tolone ; Sreyasee Das Bhattacharjee ; Huikyo Lee ; Luca Cinquini ; Shannon Schlueter ; Isaac Cho ; Wenwen Dou ; Daniel J. Crichton</i>	
ANOMALY DETECTION ON BIPARTITE GRAPHS FOR CYBER SITUATIONAL AWARENESS AND THREAT DETECTION	4741
<i>Mohammed Eslami ; George Zheng ; Hamed Eramian ; Georgiy Levchuk</i>	
EXTRACTING ROUTE PATTERNS OF VESSELS FROM AIS DATA BY USING TOPIC MODEL	4744
<i>Iwao Fujino ; Christophe Claramunt ; Abdel-Ouahab Boudraa</i>	
BIG DATA IN PSYCHOLOGY: USING WORD EMBEDDINGS TO STUDY THEORY-OF-MIND	4747
<i>Michel Généreux ; Bryor Sneiffella ; Marta Maslej</i>	
ANALYZING BIG OCEAN SCIENCE DATA WITH NEXUS	4750
<i>Frank R. Greguska ; Thomas Huang ; Brian Wilson ; Nga Quach ; Joe Jacob</i>	
TURNING BIG SPATIAL DATA INTO SMART ROUTING	4751
<i>Abdeltawab M. Hendawi ; Aqeel Rustum ; Mohamed Ali ; John Stankovic</i>	

HUMAN-CONTROLLED ITERATIVE SUBCLUSTERING ANALYSIS	4754
<i>Mauri Kaipainen ; Olli Pitkänen ; Perspicamus Ab</i>	
CONSIDERATION OF PARALLEL DATA PROCESSING OVER AN APACHE SPARK CLUSTER	4757
<i>Kasumi Kato ; Atsuko Takefusa ; Hidemoto Nakada ; Masato Oguchi</i>	
ANALYTICAL THE LARGE-SCALE COLLECTION OF DATA ON THE RESULTS OF THE GUIDES FOR FOREIGNERS VISITING JAPAN	4760
<i>Yasuko Kawahata ; Yukari Moriyama ; Shinichirou Yamada ; Mingyi Sun ; Taketo Kawamura</i>	
ISKIN SPECIALIST — A BIG DATA BASED EXPERT SYSTEM FOR DERMATOLOGY	4765
<i>Saleena Khanna ; Yuvraj Sethi ; Akash R. Nambiar</i>	
DATA ANALYTICS FOR MODELING SOIL MOISTURE PATTERNS ACROSS UNITED STATES ECOCLIMATIC DOMAINS	4768
<i>Thomas Kitson ; Paula Olaya ; Elizabeth Racca ; Michael R. Wyatt ; Mario Guevara ; Rodrigo Vargas ; Michela Taufer</i>	
GENERATING UNIFIED FAMOUS OBJECTS (UFOS) FROM THE CLASSIFIED OBJECT TABLES	4771
<i>Anusha Kola ; Harshal More ; Sean Soderman ; Michael Gubanov</i>	
ENERGY INFORMATION COLLECTION MECHANISM USING BIG DATA CORRELATION MAP	4774
<i>Tai-Yeon Ku ; Wan-Ki Park ; Hoon Choi</i>	
ANTICIPATING HUMAN ERRORS FROM PERIODIC BIG SURVEY DATA IN NUCLEAR POWER PLANTS	4777
<i>Hyun-Chul Lee ; Tong-Il Jang ; Kwangsu Moon</i>	
MAPREDUCE-BASED COMPUTATION OF AREA SKYLINE QUERY FOR SELECTING GOOD LOCATIONS IN A MAP	4779
<i>Chen Li ; Annisa ; Asif Zaman ; Yasuhiko Morimoto</i>	
DATA ANALYSIS USING HADOOP MAPREDUCE ENVIRONMENT	4783
<i>Prathyusharani Merla ; Yiheng Liang</i>	
SPATIAL-BASED TOPIC MODELLING USING WIKIDATA KNOWLEDGE BASE	4786
<i>Kwan Hui Lim ; Shanika Karunasekera ; Aaron Harwood ; Lucia Falzon</i>	
THE INFLUENCES OF DEEP-SEA VISION DATA QUALITY ON OBSERVATIONAL ANALYSIS	4789
<i>Lixin Liu ; Jun Chen</i>	
DATA-DRIVEN APPROACH TO ENSURING FAULT TOLERANCE AND EFFICIENCY OF SWARM SYSTEMS	4792
<i>Amin Majd ; Elena Troubitsyna</i>	
A SVM APPROACH FOR LIGHTPATH QOT ESTIMATION IN OPTICAL TRANSPORT NETWORKS	4795
<i>Javier Mata ; Ignacio De Miguel ; Ramón J. Durán ; Juan Carlos Aguado ; Noemí Merayo ; Lidia Ruiz ; Patricia Fernández ; Rubén M. Lorenzo ; Evaristo J. Abril</i>	
1A STUDY ON BIG DATA I/O PERFORMANCE WITH MODERN STORAGE SYSTEMS	4798
<i>Kenji Nakashima ; Joichiro Kon ; Saneyasu Yamaguchi ; Gil Jae Lee ; José Fortes</i>	
BIOFEEDBACK EEG DATA INTEGRATION AND VISUALIZATION ANALYTICS FOR ENDURANCE EXERCISE PRACTICES: DATA INTEGRATION AND VISUALIZATION ANALYTICS OF BIOFEEDBACK EEG	4800
<i>Monika Nawrocka ; Marcin Lukowski</i>	
A PERFORMANCE EVALUATION OF APACHE KAFKA IN SUPPORT OF BIG DATA STREAMING APPLICATIONS	4803
<i>Paul Le Noac'H ; Alexandru Costan ; Luc Bougé</i>	
HYBRID.JSON: HIGH-VELOCITY PARALLEL IN-MEMORY POLYSTORE JSON INGEST	4807
<i>Steven Ortiz ; Caner Enbatan ; Maksim Podkorytov ; Dylan Soderman ; Michael Gubanov</i>	
USING BI-PARTITE GRAPHS TO CLUSTER COMPLEX NETWORKS	4810
<i>Kaine Black ; Monica Wachowicz ; Alec Parise</i>	
ART-2B: ADAPTED ART-2A FOR LARGE SCALE DATA CLUSTERING ON PM2.5 MASS SPECTRA	4813
<i>Nat Pavasant ; Hiroshi Furutani ; Masayuki Numao ; Ken-Ichi Fukui</i>	
AUTOMATIC KEYWORD EXTRACTION: AN ENSEMBLE METHOD	4816
<i>Tayfun Pay ; Stephen Lucci</i>	
THE CASE FOR GRAPH-BASED RECOMMENDATIONS	4819
<i>Iulia Popescu ; Kurt Portelli ; Christos Anagnostopoulos ; Nikos Ntarmos</i>	

BASELINES FOR DEMOGRAPHIC INFERENCE ON A NEW GOLD STANDARD TWITTER CORPUS.....	4822
<i>Jason Radford ; Luke Horgan ; David Lazer</i>	
PILOTING A THEORY-BASED APPROACH TO INFERRING GENDER IN BIG DATA	4824
<i>Jason Radford</i>	
PRIVACY-PRESERVING OUTSOURCED COLLABORATIVE FREQUENT ITEMSET MINING IN THE CLOUD.....	4827
<i>Bharath K. Samanthula</i>	
A STUDY ON INTERPRETABILITY OF DECISION OF MACHINE LEARNING.....	4830
<i>Shohei Shirataki ; Saneyasu Yamaguchi</i>	
HYBRID.MEDIA: HIGH VELOCITY VIDEO INGESTION IN AN IN-MEMORY SCALABLE ANALYTICAL POLYSTORE.....	4832
<i>Mark Simmons ; Daniel Armstrong ; Dylan Soderman ; Michael Gubanov</i>	
EOS: A MULTILINGUAL TEXT ARCHIVE OF INTERNATIONAL NEWSPAPER & BLOG ARTICLES.....	4835
<i>Lisa Singh ; Raghu Pemmaraju</i>	
APPLICATION SPECIFIC TRAFFIC CONTROL IN LARGE-SCALE DISASTERS.....	4838
<i>Tsumugi Tairaku ; Akihiro Nakao ; Saneyasu Yamaguchi ; Masato Oguchi</i>	
ROAD SAFETY ESTIMATION UTILIZING BIG AND HETEROGENEOUS VEHICLE RECORDER DATA	4841
<i>Masashi Toyoda ; Daisaku Yokoyama ; Junpei Komiyama ; Masahiko Itoh</i>	
REAL TIME ANALYTICS — STATE OF THE ART: POTENTIALS AND LIMITATIONS IN THE SMART FACTORY	4843
<i>Sebastian Trinks ; Carsten Felden</i>	
MCMALLOC: A SCALABLE MEMORY ALLOCATOR FOR MULTITHREADED APPLICATIONS ON A MANY-CORE SHARED-MEMORY MACHINE	4846
<i>Akira Umayabara ; Hayato Yamana</i>	
SCALABLE SPAM CLASSIFIER FOR WEB TABLES.....	4849
<i>Santiago Villasenor ; Tom Nguyen ; Anusha Kola ; Sean Soderman ; Michael Gubanov</i>	
ACCURATE SIGNAL TIMING FROM HIGH FREQUENCY STREAMING DATA.....	4852
<i>Jonathan Wang ; Kesheng Wu ; Alex Sim ; Seongwook Hwangbo</i>	
UNDERSTANDING THE IMPACT OF SAMPLING AND NOISE ON DETECTING EVENTS USING TWITTER	4855
<i>Yifang Wei ; Lisa Singh</i>	
ATTRIBUTE-BASED PROXY RE-ENCRYPTION METHOD FOR REVOCATION IN CLOUD DATA STORAGE	4858
<i>Yoshiko Yasumura ; Hiroki Imabayashi ; Hayato Yamana</i>	
TOWARDS CONSTRUCTING A DRIVER MANAGEMENT SYSTEM BASED ON LARGE-SCALE DRIVING OPERATION RECORDS	4861
<i>Daisaku Yokoyama ; Masashi Toyoda</i>	
PROPOSAL OF CLASSIFICATION METHOD OF BUS OPERATION STATES USING SENSOR DATA.....	4863
<i>Takuya Yonezawa ; Ismail Arai ; Toyokazu Akiyama ; Kazutoshi Fujikawa</i>	
UNDERSTANDING A MODERATING EFFECT OF PHYSICIANS' ENDORSEMENT TO ONLINE WORKLOAD: AN EMPIRICAL STUDY IN ONLINE HEALTH-CARE COMMUNITIES.....	4866
<i>Haiyan Yu ; Kun Xiang ; Jiang Yu</i>	
TOWARDS AUTOMATIC INFRASTRUCTURE PROVISIONING FOR HIGHLY DYNAMIC STREAMING APPLICATIONS.....	4869
<i>Philipp Zehnder ; Dominik Riemer</i>	
PERSONALIZED SEARCH WITH EDITABLE PROFILES.....	4872
<i>Binyam A. Zemedu ; Byron J. Gao</i>	
DISCOVERING THE INTERDISCIPLINARY NATURE OF BIG DATA RESEARCH	4875
<i>Yin Zhang ; Jiming Hu</i>	
BIG DATA SYSTEM FOR INFORMATION AGGREGATION AND MODEL COMPARISON FOR PRECISION MEDICINE.....	4878
<i>Ziwei Zhu ; Weijia Xu ; Wei He</i>	
Author Index	