

2017 IEEE International Conference on Information Reuse and Integration (IRI 2017)

**San Diego, California, USA
4-6 August 2017**



**IEEE Catalog Number: CFP17IRI-POD
ISBN: 978-1-5386-1563-8**

**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:	CFP17IRI-POD
ISBN (Print-On-Demand):	978-1-5386-1563-8
ISBN (Online):	978-1-5386-1562-1

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

2017 IEEE International Conference on Information Reuse and Integration

IRI 2017

Table of Contents

Foreword.....	xiii
Message from Program Co-Chairs.....	xiv
Conference Organization.....	xvi
IRI 2017 Program Committee.....	xviii
IRI 2017 Steering Committee.....	xx
DIM 2017 Workshop Organization.....	xxi
EM-RITE 2017 Workshop Organization.....	xxii
FMi 2017 Workshop Organization.....	xxiii
IICPS 2017 Workshop Organization.....	xxiv
SMA2 2017 Workshop Organization.....	xxv
Keynote Talks.....	xxvi

Session A11: Security and Privacy

Towards Managing Complex Data Sharing Policies with the Min Mask Sketch	1
<i>Stephen Smart and Christan Grant</i>	
Cyber-Secure UAV Communications Using Heuristically Inferred Stochastic Grammars and Hard Real-Time Adaptive Waveform Synthesis and Evolution	9
<i>Stuart H. Rubin, William K. Grefe, Thouraya Bouabana-Tebibel, Shu-Ching Chen, Mei-Ling Shyu, and Kenneth S. Simonsen</i>	
Access Control Models for Virtual Object Communication in Cloud-Enabled IoT	16
<i>Asma Alshehri and Ravi Sandhu</i>	
Malware Collection and Analysis	26
<i>Ramkumar Paranthaman and Bhavani Thuraisingham</i>	

Session A12: Modeling and Mining

Unsupervised Terminological Ontology Learning Based on Hierarchical Topic Modeling	32
<i>Xiaofeng Zhu, Diego Klabjan, and Patrick N. Bless</i>	
Extracting Knowledge from Technical Reports for the Valuation of West Texas Intermediate Crude Oil Futures	42
<i>Joseph D. Prusa, Ryan Sagul, Taghi M. Khoshgoftaar, and Michael Sterling</i>	
Business Objects plus (BO+): An Approach to Enhance Service Reuse and Integration in Cross-Domain SOA Compounds	49
<i>Jan Königsberger and Bernhard Mitschang</i>	
On Using Disparate Scholarly Data to Identify Potential Members for Interdisciplinary Research Groups	59
<i>Francisco Osuna, Monika Akbar, and Ann Q. Gates</i>	

IICPS 2017: International Workshop on Information Integration in Cyber Physical Systems

Traffic Surveillance System for Bridge Vibration Analysis	69
<i>Takaya Kawakatsu, Akira Kakitani, Kenro Aihara, Atsuhiko Takasu, and Jun Adachi</i>	
Blockchain-Assisted Information Distribution for the Internet of Things	75
<i>George C. Polyzos and Nikos Fotiou</i>	
A Collaborative Robotic Cyber Physical System for Surgery Applications	79
<i>Daniela D'Auria and Fabio Persia</i>	
Active Dependency Mapping: A Data-Driven Approach to Mapping Dependencies in Distributed Systems	84
<i>Alexia Schulz, Michael Kotson, Chad Meiners, Timothy Meunier, David O'Gwynn, Pierre Trepagnier, and David Weller-Fahy</i>	

Session A2: Decision Support

Automated Dynamic Negotiation over Environmental Issues	92
<i>Faezeh Eshragh, Mozhdeh Shahbazi, and Behrouz Far</i>	
A Scalable Spark-Based Fault Diagnosis Platform for Gearbox Fault Diagnosis in Wind Farms	100
<i>Maryam Bahojb Imani, Mehrdad Heydarzadeh, Latifur Khan, and Mehrdad Nourani</i>	
A Cyber-Provenance Infrastructure for Sensor-Based Data-Intensive Applications	108
<i>Elisa Bertino and Murat Kantarcioglu</i>	

A Novel Approach to Neural Network Based Intelligent Systems for Emotion Recognition in Audio Signals	115
<i>Abdelniser Mooman and Stephen Powers</i>	

Session APP1: Applications: Detection

Constructing Financial Sentiment Lexicons by Integrating Textual and Time-Series Data	121
<i>Kazuhiro Seki and Masahiko Shibamoto</i>	
Towards Using KDD for an Interactive Software Architecture Extraction	127
<i>Mira Abboud, Hala Naja, Mohamad Dbouk, and Mourad Oussalah</i>	
Fresh-Phish: A Framework for Auto-Detection of Phishing Websites	137
<i>Hossein Shirazi, Kyle Haefner, and Indrakshi Ray</i>	
Mining Social Media Data Using Topological Data Analysis	144
<i>Khaled Almgren, Minkyu Kim, and Jeongkyu Lee</i>	
User Behavior Anomaly Detection for Application Layer DDoS Attacks	154
<i>Maryam M. Najafabadi, Taghi M. Khoshgoftaar, Chad Calvert, and Clifford Kemp</i>	
Species Distribution Modeling with Scalability: The Case Study of P-GARP, a Parallel Genetic Algorithm for Rule-Set Production	162
<i>Fabiana Santana, César Alberto Bravo Pariente, and Antonio Mauro Saraiva</i>	

DIM 2017: International Workshop on Data Integration and Mining

A Hybrid Machine Learning Method for Finding Depression Related Publications by Eliminating Outlier Publications	171
<i>Zhibo Wang, Long Ma, and Yanqing Zhang</i>	
Addressing Forest Management Challenges by Refining Tree Cover Type Classification with Machine Learning Models	177
<i>Duncan Macmichael and Dong Si</i>	
An Architecture for the Internet of Things and the Use of Big Data Techniques in the Analysis of Carbon Monoxide	184
<i>Marco Aurelio Borges, Paulo Batista Lopes, Leandro A. Silva, Massaki de O. Igarashi, and Gabriel Melo F. Correia</i>	
A Software Solution for Preventing Alzheimer's Disease Based on MapReduce Framework	192
<i>Manal Zettam, Jalal Laassiri, and Nouredine Enneya</i>	

Session B11: Diagnosis and Prediction

FA-MCADF: Feature Affinity Based Multiple Correspondence Analysis and Decision Fusion Framework for Disaster Information Management	198
<i>Haiman Tian, Shu-Ching Chen, Stuart H. Rubin, and William K. Grefe</i>	
Link Prediction and Hybrid Strategies for Updating a Social Graph Snapshot via a Limited API	207
<i>Norases Vesdapunt and Hector Garcia-Molina</i>	
Detecting Small Signs from Large Images	217
<i>Zibo Meng, Xiaochuan Fan, Xin Chen, Min Chen, and Yan Tong</i>	
EARRING: Efficient Authentication of Outsourced Record Matching	225
<i>Boxiang Dong and Hui (Wendy) Wang</i>	

Session B12: Knowledge Representation and Management

GFEL: Generalized Feature Embedding Learning Using Weighted Instance Matching	235
<i>Eric Golinko and Xingquan Zhu</i>	
Accelerated Sparse Representation for Human Activity Recognition	245
<i>Long Cheng, Yani Guan, Kecheng Zhu, Yiyang Li, and Ruokun Xu</i>	
A Multi-stage Approach to Personalized Course Selection and Scheduling	253
<i>Tyler Morrow, Ali R. Hurson, and Sahra Sedigh Sarvestani</i>	
Colored Petri Net Representation of Logical and Decisive Passing Algorithm for Humanoid Soccer Robots	263
<i>Kieutran Theresa Pham, Chelsea Cantone, and Seung-Yun Kim</i>	

FMI 2017: International Workshop on Formal Methods Integration I

Upgrade of Highly Available Systems: Formal Methods at the Rescue	270
<i>Oussama Jebbar, Ferhat Khendek, and Maria Toeroe</i>	
Probabilistic Analysis Applied to Cleaning Robots	275
<i>Rafael Pereira de Araújo, Alexandre Cabral Mota, and Sidney de Carvalho Nogueira</i>	
A Symbolic Rewriting Semantics of the COMPASS Modeling Language	283
<i>Musab A. Alturki</i>	
A Phenomenal Basis for Hybrid Modelling	291
<i>Jon G. Hall, Lucia Rapanotti, and Georgi Markov</i>	

Session APP2: Applications: Social Media and Users

Analysis and Modeling of Unified User Interest	298
<i>Sampath Jayarathna and Frank Shipman</i>	
Analysis of Performance Variability in Public Cloud Computing	308
<i>Fabiana Santana</i>	
Model-Oriented Web Service Implementations Compared to Traditional Web Services	315
<i>Thiago Gottardi and Rosana Teresinha Vaccare Braga</i>	

Session APP3: Applications: AI and Machine Learning

Modified Convolution Neural Network for Highly Effective Parallel Processing	325
<i>Sang-Soo Park, Jung-Hyun Hong, and Ki-Seok Chung</i>	
An Automated Approach for Information and Referral of Social Services Using Machine Learning	332
<i>Madhav Sharan, Nithin Krishna Ottilingam, Chris A. Mattmann, Karanjeet Singh, Maribel Marin, Amy Latzer, Colin Foon, and Umesh Handore</i>	
Analysis of Transfer Learning Performance Measures	338
<i>Karl R. Weiss and Taghi M. Khoshgoftaar</i>	

Session B2: Learning

Comparison of Visual Datasets for Machine Learning	346
<i>Kent Gauen, Ryan Dailey, John Laiman, Yuxiang Zi, Nirmal Asokan, Yung-Hsiang Lu, George K. Thiruvathukal, Mei-Ling Shyu, and Shu-Ching Chen</i>	
AAFA: Associative Affinity Factor Analysis for Bot Detection and Stance Classification in Twitter	356
<i>Saad Sadiq, Yilin Yan, Asia Taylor, Mei-Ling Shyu, Shu-Ching Chen, and Daniel Feaster</i>	
Using Weather and Playing Surface to Predict the Occurrence of Injury in Major League Soccer Games: A Case Study	366
<i>Sara Landset, Michael F. Bergeron, and Taghi M. Khoshgoftaar</i>	
Documentation Reuse: Managing Similar Documents	372
<i>Stanislaw Jarzabek and Daniel Dan</i>	

EM-RITE 2017: 6th International Workshop on Empirical Methods for Recognizing Inference in Text

Visualization Analysis of Medical Resource Utilization for Long-Term Care Services	376
<i>Kuo-Chung Chu and Peng-Hua Jiang</i>	
Deep Learning for Sentiment Analysis on Google Play Consumer Review	382
<i>Min-Yuh Day and Yue-Da Lin</i>	
Heuristics-Based Schema Extraction for Deep Web Query Interfaces	389
<i>Chichang Jou and Yucheng Cheng</i>	
Constructing Hierarchical Product Categories for E-Commerce by Word Embedding and Clustering	397
<i>Yi-Hsiang Hsieh, Shih-Hung Wu, Liang-Pu Chen, and Ping-Che Yang</i>	

FMi 2017: International Workshop on Formal Methods Integration II

On the Synthesis of Guaranteed-Quality Plans for Robot Fleets in Logistics Scenarios via Optimization Modulo Theories	403
<i>Francesco Leofante, Erika Ábrahám, Tim Niemueller, Gerhard Lakemeyer, and Armando Tacchella</i>	
Performance Analysis of Apache Storm Applications Using Stochastic Petri Nets	411
<i>Jose-Ignacio Requeno, Jose Merseguer, and Simona Bernardi</i>	
On the Verification of UML State Machine Diagrams to Colored Petri Nets Transformation Using Isabelle/HOL	419
<i>Said Meghzili, Allaoua Chaoui, Martin Strecker, and Elhillali Kerkouche</i>	

Industry/Gov Session & Workshop

The Challenges and Solutions for Building an Integrated Video Analytics Platform	427
<i>George Mathew</i>	
Predicting the Users' Clickstreams Using Time Series Representation, Symbolic Sequences, and Deep Learning: Application on Job Offers Recommendation Tasks	436
<i>Sidahmed Benabderrahmane, Nedra Mellouli, and Myriam Lamolle</i>	
ULTR-CTR: Fast Page Grouping Using URL Truncation for Real-Time Click Through Rate Estimation	444
<i>Hui Liu, Xingquan Zhu, Kristopher Kalish, and Jeremy Kayne</i>	

Session C1: Social Media and Users

Enhancing Yelp Data with Deep Learning and Information Reuse	452
<i>Roberto Camacho Barranco, Laura M. Rodriguez, Rebecca Urbina, and M. Shahriar Hossain</i>	
Estimating the Prevalence of Religious Content in Intelligent Design Social Media	462
<i>George D. Montañez</i>	
Infusing Latent User-Concerns from User Reviews into Collaborative Filtering	471
<i>Ligaj Pradhan, Chengcui Zhang, and Steven Bethard</i>	

SMA2 2017: First International Workshop on Social Media Networks Analytics and Applications I

Predicting Movie Market Revenue Using Social Media Data	478
<i>Steve Shim and Mohammad Pourhomayoun</i>	
Discovering Mobility Patterns of Instagram Users through Process Mining Techniques	485
<i>Claudia Diamantini, Laura Genga, Fabrizio Marozzo, Domenico Potena, and Paolo Trunfio</i>	
Modelling Multimedia Social Networks Using Semantically Labelled Graphs	493
<i>Enrico G. Caldarola and Antonio M. Rinaldi</i>	
Influence Analysis in Online Social Networks Using Hypergraphs	501
<i>Flora Amato, Francesco di Lillo, Vincenzo Moscato, Antonio Picariello, and Giancarlo Sperli</i>	

Session APP4: Applications: Information Reuse

Ensemble Maximum Entropy Classification and Linear Regression for Author Age Prediction	509
<i>Joey Hong, Chris A. Mattmann, and Paul Ramirez</i>	
Toward Semantic Search for the Biogeochemical Literature	517
<i>Joshua D. Eisenberg, Deya Banisakher, Maria Presa, Kalli Unthank, Mark A. Finlayson, Rene Price, and Shu-Ching Chen</i>	
Entropy in Design Phase: A Higraph-Based Model Approach	526
<i>Hycham Aboutaleb and Bruno Monsuez</i>	

Session APP5: Applications: Data Analytics

An Effective LmRMR for Financial Variable Selection and its Applications	535
<i>Sara Aghakhani, Reda Alhadj, Jon Rokne, and Philip Chang</i>	
Learning Profiles in Duplicate Question Detection	544
<i>Chakaveh Saedi, João Rodrigues, João Silva, António Branco, and Vladislav Maraev</i>	
Modernizing Analytics for Melanoma with a Large-Scale Research Dataset	551
<i>Aaron N. Richter and Taghi M. Khoshgoftaar</i>	

Session C2: Classification and Processing

Estimating Outlier Score Probabilities	559
<i>Richard A. Bauder and Taghi M. Khoshgoftaar</i>	
Schemaless Join for Result Set Preferences	569
<i>Chuancong Gao, Jian Pei, Jiannan Wang, and Yi Chang</i>	
Medical Provider Specialty Predictions for the Detection of Anomalous Medicare Insurance Claims	579
<i>Matthew Herland, Richard A. Bauder, and Taghi M. Khoshgoftaar</i>	
Knowledge Amplification through Randomization for Scheduling Systems	589
<i>Thouraya Bouabana-Tebibel, Stuart H. Rubin, Miled B. Bentaiba, Abdelghani Allaoua, and Ahcene Boumhand</i>	

SMA2 2017: First International Workshop on Social Media Networks Analytics and Applications II

Effective Information Spreading in Social Networks	599
<i>Nunziato Cassavia, Elio Masciari, and Domenico Saccà</i>	
Measuring Polarization in Twitter Enabled in Online Political Conversation: The Case of 2016 US Presidential Election	607
<i>Simonetta Primario, Dario Borrelli, Luca Iandoli, Giuseppe Zollo, and Carlo Lipizzi</i>	
A Survey of Online Social Networks: Challenges and Opportunities	614
<i>Fabio Persia and Daniela D'Auria</i>	
Author Index	621