

2016 IEEE 17th International Conference on Information Reuse and Integration (IRI 2016)

**Pittsburgh, Pennsylvania, USA
28-30 July 2016**



**IEEE Catalog Number: CFP16IRI-POD
ISBN: 978-1-5090-3208-2**

**Copyright © 2016 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 publication is a representation of what appears in the IEEE Digital Libraries. Some format issues inherent in the e-media version may also appear in this print version.***

| | |
|-------------------------|-------------------|
| IEEE Catalog Number: | CFP16IRI-POD |
| ISBN (Print-On-Demand): | 978-1-5090-3208-2 |
| ISBN (Online): | 978-1-5090-3207-5 |

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

2016 IEEE 17th International Conference on Information Reuse and Integration

IRI 2016

Table of Contents

| | |
|---|-------|
| Foreword from the General Chairs..... | xiii |
| Message from the Program Co-Chairs..... | xiv |
| Organizing Committee..... | xvi |
| Steering Committee..... | xvii |
| Keynote Abstracts..... | xviii |

Keynote Paper

| | |
|---|---|
| A Data Driven Approach for the Science of Cyber Security: Challenges and Directions | 1 |
| <i>Bhavani Thuraisingham, Murat Kantarcioglu, Kevin Hamlen, Latifur Khan, Tim Finin, Anupam Joshi, Tim Oates, and Elisa Bertino</i> | |

Session A11: Security and Privacy

| | |
|---|----|
| A Novel Method for Fraudulent Medicare Claims Detection from Expected Payment Deviations (Application Paper) | 11 |
| <i>Richard A. Bauder and Taghi M. Khoshgoftaar</i> | |
| Anomaly Detection Techniques for Database Protection Against Insider Threats (Invited Paper) | 20 |
| <i>Asmaa Sallam, Qian Xiao, Elisa Bertino, and Daren Fadolalkarim</i> | |
| Improving the Utility of Anonymized Datasets through Dynamic Evaluation of Generalization Hierarchies | 30 |
| <i>Vanessa Ayala-Rivera, Thomas Cerqueus, Liam Murphy, and Christina Thorpe</i> | |
| A Risk-Benefit Driven Architecture for Personal Data Release (Invited Paper) | 40 |
| <i>B. C. Singh, B. Carminati, and E. Ferrari</i> | |

Session A12: Reuse and Modeling

| | |
|---|----|
| Reusable Meta-Models for Crowdsourcing Driven Elastic Systems (Invited Paper) | 50 |
| <i>Julian Jarrett and M. Brian Blake</i> | |
| Modeling Integration and Reuse of Heterogeneous Terminologies in Faceted Browsing Systems | 58 |
| <i>Daniel R. Harris</i> | |
| Discover Tipping Users for Cross Network Influencing (Invited Paper) | 67 |
| <i>Qianyi Zhan, Jiawei Zhang, Philip S. Yu, Sherry Emery, and Junyuan Xie</i> | |

Session A13: Spatial and Mobile Systems

| | |
|---|-----|
| A Machine Learning Approach to Improve the Accuracy of GPS-Based Map-Matching Algorithms (Invited Paper) | 77 |
| <i>Mahdi Hashemi and Hassan A. Karimi</i> | |
| An Automatic Approach for Discovering and Geocoding Locations in Domain-Specific Web Data (Application Paper) | 87 |
| <i>Chris A. Mattmann and Madhav Sharan</i> | |
| GeoTeCS: Exploiting Geographical, Temporal, Categorical and Social Aspects for Personalized POI Recommendation (Invited Paper) | 94 |
| <i>Ramesh Baral, Dingding Wang, Tao Li, and Shu-Ching Chen</i> | |
| Predicting Students' Behavioral Patterns in University Networks for Efficient Bandwidth Allocation: A Hybrid Data Mining Method (Application Paper) | 102 |
| <i>Elham Akhond Zadeh Noughabi, Behrouz H. Far, and Bijan Raahemi</i> | |

IEEE International Workshop on Formal Methods Integration (FMI)

| | |
|--|-----|
| Modular Verification of Termination and Execution Time Bounds Using Separation Logic | 110 |
| <i>Jafar Hamin and Bart Jacobs</i> | |
| Modelling Complex Timing Requirements with Refinement | 118 |
| <i>Gintautas Sulskus, Michael Poppleton, and Abdolbaghi Rezazadeh</i> | |
| Spekl: A Layered System for Specification Authoring, Sharing, and Usage | 126 |
| <i>John L. Singleton and Gary T. Leavens</i> | |
| A Formal Language for Writing Contracts | 134 |
| <i>William M. Farmer and Qian Hu</i> | |

Session A21: Classification & Processing

| | |
|--|-----|
| Monitoring Classification Blindspots to Detect Drifts from Unlabeled Data | 142 |
| <i>Tegjyot Singh Sethi, Mehmed Kantardzic, and Elaheh Arabmakki</i> | |
| Designing a Testing Framework for Transfer Learning Algorithms (Application Paper) | 152 |
| <i>Karl R. Weiss, Taghi M. Khoshgoftaar, and Oneeb Rehman</i> | |
| Cross-Domain Sentiment Analysis: An Empirical Investigation | 160 |
| <i>Brian Heredia, Taghi M. Khoshgoftaar, Joseph Prusa, and Michael Crawford</i> | |

Session A22: Learning

| | |
|---|-----|
| Classifying Questions into Fine-Grained Categories Using Topic Enriching | 166 |
| <i>Zhao Lu, Yu-Ru Lin, Qing Zhang, and Mengwei Chen</i> | |
| Clustering Web Pages Based on Structure and Style Similarity (Application Paper) | 175 |
| <i>Thamme Gowda and Chris A. Mattmann</i> | |
| Automatic Label Data Abstraction Based on Information Entropy (Application Paper) | 181 |
| <i>Fumiya Kudo, Norihiko Moriwaki, Tomoaki Akitomi, Susumu Serita, and Yu Kitano</i> | |
| Algorithmic Decision Support for Personalized Education (Invited Paper) | 188 |
| <i>Tyler Morrow, Sahra Sedigh Sarvestani, and Ali R. Hurson</i> | |

Session A23: Database and Cloud Systems

| | |
|--|-----|
| Employing Graph Databases as a Standardization Model towards Addressing Heterogeneity | 198 |
| <i>Dippy Aggarwal and Karen C. Davis</i> | |
| Query-Driven Sampling for Collective Entity Resolution | 208 |
| <i>Christan Grant, Daisy Zhe Wang, and Michael Wick</i> | |
| An Extensible Software Platform for Cloud-Based Decision Support and Automation in Precision Agriculture | 218 |
| <i>Li Tan, Hongfei Hou, and Qin Zhang</i> | |
| Cloud-Based Web Services for Endangered Language Analysis (Invited Paper) | 226 |
| <i>Ashik Samad, Kevin Wu, and Min Chen</i> | |

IEEE International Workshop on Formal Methods Integration (FMI)

| | |
|--|-----|
| Toward Bringing Distributed System Design upon Rigorous Footing | 233 |
| <i>Mohammed S. Al-Mahfoudh, Ganesh Gopalakrishnan, and Ryan Stutsman</i> | |

| | |
|---|-----|
| Support for Model Checking Z Specifications | 241 |
| <i>Maria Ulfah Siregar</i> | |

| | |
|---|-----|
| BHive: Towards Behaviour-Driven Development Supported by B-Method | 249 |
| <i>J. Carter and W. B. Gardner</i> | |

Poster Session

| | |
|--|-----|
| Ensemble Classifier for Imbalanced Streaming Data Using Partial Labeling | 257 |
| <i>Elaheh Arabmakki, Mehmed Kantardzic, and Tegjyot Singh Sethi</i> | |

| | |
|--|-----|
| Reuse of WiFi Information for Indoor Monitoring of the Elderly | 261 |
| <i>Xuan Xie, Hongze Xu, Guang Yang, Zhi-Hong Mao, Wenyan Jia, and Mingui Sun</i> | |

| | |
|--|-----|
| Development of Efficient Obstacle Avoidance for a Mobile Robot Using Fuzzy Petri Nets | 265 |
| <i>Philip D. Baldoni, Yilin Yang, and Seung-Yun Kim</i> | |

| | |
|--|-----|
| Analysis of Soccer Robot Behaviors Using Time Petri Nets | 270 |
| <i>Daniel Ponsini, Yilin Yang, and Seung-Yun Kim</i> | |

| | |
|--|-----|
| Datom Continued: Towards Multi-Objective Optimization of Data Management Entities | 275 |
| <i>Verena Kantere</i> | |

| | |
|--|-----|
| Intrusion Detection Using a Multiple-Detector Set Artificial Immune System | 283 |
| <i>James Brown, Mohd Anwar, and Gerry Dozier</i> | |

Session B11: Access Control and Trust

| | |
|---|-----|
| PETRA: Personalized Trust Management Architecture (Application Paper) | 287 |
| <i>Hisham Salah and Mohamed Eltoweissy</i> | |

| | |
|--|-----|
| Object-to-Object Relationship-Based Access Control: Model and Multi-Cloud Demonstration (Invited Paper) | 297 |
| <i>Tahmina Ahmed, Farhan Patwa, and Ravi Sandhu</i> | |

| | |
|---|-----|
| An In-Depth Study of Peer Influence on Collective Decision Making for Multi-party Access Control (Invited Paper) | 305 |
| <i>Alpana Tyagi, Anna Squicciarini, Sarah Rajtmajer, and Christopher Griffin</i> | |

| | |
|--|-----|
| General Methods for Access Control Policy Verification (Application Paper) | 315 |
| <i>Vincent C. Hu and D. Richard Kuhn</i> | |

Session B12: Infrastructure Resilience and Disaster Response

| | |
|--|-----|
| A Scalable and Automatic Validation Process for Florida Public Hurricane Loss Model (Invited Paper) | 324 |
| <i>Haiman Tian, Hsin-Yu Ha, Samira Pouyanfar, Yilin Yan, Sheng Guan, Shu-Ching Chen, Mei-Ling Shyu, and Shahid Hamid</i> | |
| Integrating Control and Fault-Tolerant Wireless Network Design for Small Modular Nuclear Reactors (Invited Paper) | 332 |
| <i>Wenchen Wang, Christopher D'Angelo, Daniel Mosse, and Daniel Cole</i> | |
| Identification of Power Line Outages Based on PMU Measurements and Sparse Overcomplete Representation | 343 |
| <i>Long Cheng, Chenyu You, and Liguo Chen</i> | |
| Selective Mapping for Real-Time Video Transport over Wireless LAN | 350 |
| <i>Hongli Luo</i> | |

IEEE International Workshop on Data Integration and Mining I (DIM)

| | |
|--|-----|
| Generating Headline Candidates for News Articles | 355 |
| <i>Naoya Okumura and Takao Miura</i> | |
| Process Discovery Using Classification Tree Hidden Semi-Markov Model | 361 |
| <i>Yihuang Kang and Vladimir Zadorozhny</i> | |

IEEE Workshop on Information Reuse and Integration in Health Informatics (IRI-HI)

| | |
|--|-----|
| Predicting Cancer Relapse with Clinical Data: A Survey of Current Techniques | 369 |
| <i>Aaron N. Richter and Taghi M. Khoshgoftaar</i> | |
| Investigating the Variation of Ensemble Size on Bagging-Based Classifier Performance in Imbalanced Bioinformatics Datasets | 377 |
| <i>Alireza Fazelpour, Taghi M. Khoshgoftaar, David J. Dittman, and Amri Naplitano</i> | |

Session B21: Information Processing and Reuse

| | |
|---|-----|
| An Approach to Ontology Integration for Ontology Reuse | 384 |
| <i>Enrico G. Caldarola and Antonio M. Rinaldi</i> | |
| Information Fusion Using Characteristic Linear System in Multi-robot Search and Rescue Task | 394 |
| <i>Pei-Ju Lee and Vladimir Zadorozhny</i> | |
| Bloom Filter Tree for Fast Search and Synchronization of Tree-Structured Data | 401 |
| <i>Mengyu Wang and Ying Zhu</i> | |

| | |
|---|-----|
| Designing a Better Data Representation for Deep Neural Networks and Text Classification | 411 |
| <i>Joseph D. Prusa and Taghi M. Khoshgoftaar</i> | |

Session B22: Data Storage and Analytics

| | |
|---|-----|
| Data Analytics Using Ontologies of Management Theories: Towards Implementing "From Theory to Practice" (Application Paper) | 417 |
| <i>Henry M. Kim, Jackie Ho Nam Cheung, Marek Laskowski, and Iryna Gel</i> | |
| A Semantically-Enabled SOA Governance Repository (Application Paper) | 423 |
| <i>Jan Königsberger and Bernhard Mitschang</i> | |
| Statistical Measurement and Analysis on How the Late-Life Function & Disability Instrument Enhances the Frailty Assessment Compared to the National Standards Used on Transcatheter Aortic Valve Patients (Application Paper) | 433 |
| <i>Yunkai Liu, A. Kate MacPhedran, and Yu Luo</i> | |
| Domain Knowledge Assisted Data Processing for Florida Public Hurricane Loss Model (Invited Paper) | 441 |
| <i>Yilin Yan, Samira Pouyanfar, Haiman Tian, Sheng Guan, Hsin-Yu Ha, Shu-Ching Chen, Mei-Ling Shyu, and Shahid Hamid</i> | |

IEEE International Workshop on Data Integration and Mining II (DIM)

| | |
|---|-----|
| Rapid Prototyping of a Text Mining Application for Cryptocurrency Market Intelligence | 448 |
| <i>Marek Laskowski and Henry M. Kim</i> | |
| Discovering News Frames: Exploring Text, Content, and Concepts in Online News Sources to Address Water Insecurity in the Southwest Region | 454 |
| <i>Loretta H. Cheeks, Tracy L. Stepien, and Dara M. Wald</i> | |
| A Topic-Independent Hybrid Approach for Sentiment Analysis of Chinese Microblog | 463 |
| <i>Ping Han, Shan Li, and Yunfei Jia</i> | |
| Fast Document Indexing Using Aho-Corasick State Machine | 469 |
| <i>Ramanpreet Singh and Gurpreet Singh Mohaar</i> | |
| Recommender System Framework for Academic Choices: Personality Based Recommendation Engine (PBRE) | 476 |
| <i>Muhammad Fahim Uddin, Soumita Banerjee, and Jeongkyu Lee</i> | |

IEEE International Workshop on Information Integration in Cyber Physical Systems (IICPS)

| | |
|---|-----|
| Comparison of Real-Time, Mobile Data Bus Architectures | 484 |
| <i>Paul M. Simon and Scott R. Graham</i> | |
| An Event Detection Framework Supported by a Smart Graphical User Interface | 490 |
| <i>Fabio Persia and Sven Helmer</i> | |
| Tracked Robot over a Slope Path: Dynamic Stability Control | 496 |
| <i>Daniela D'Auria, Gianluca Ristorto, Gallo Raimondo, and Fabrizio Mazzetto</i> | |
| Anomaly Detection in Electrical Substation Circuits via Unsupervised Machine Learning | 500 |
| <i>Alfonso Valdes, Richard Macwan, and Matthew Backes</i> | |

IEEE International Workshop on Empirical Methods for Recognizing Inference in Text (EMRITE)

| | |
|--|-----|
| Learning Textual Entailment Classification from a Chinese RITE Dataset Specialized for Linguistic Phenomena | 506 |
| <i>Chi-Ting Liu, Shao-Heng Chen, and Chuan-Jie Lin</i> | |
| A Study on Machine Learning for Imbalanced Datasets with Answer Validation of Question Answering | 513 |
| <i>Min-Yuh Day and Cheng-Chia Tsai</i> | |
| Principle-Based Approach for Semi-Automatic Construction of a Restaurant Question Answering System from Limited Datasets | 520 |
| <i>Ting-Hao Yang, Yu-Lun Hsieh, You-Shan Chung, Cheng-Wei Shih, Shih-Hung Liu, Yung-Chun Chang, and Wen-Lian Hsu</i> | |
| Improving the Selection Error Recognition in a Chinese Grammar Error Detection System | 525 |
| <i>Po-Lin Chen, Shih-Hung Wu, Liang-Pu Chen, and Ping-Che Yang</i> | |

Session C11: Semantics for Reuse and Integration

| | |
|--|-----|
| Reusing Software Libraries Using Semantic Graphs | 531 |
| <i>Awmy Alnusair, Majdi Rawashdeh, Mohammed F. Alhamid, M. Anwar Hossain, and Ghulam Muhammad</i> | |
| Knowledge Induction Based on Randomization in Case-Based Reasoning | 541 |
| <i>Thouraya Bouabana-Tebibel, Stuart H. Rubin, Asmaa Chebba, Sofiane Bediar, and Syphax Iskounen</i> | |
| Semantic Interpretation of Structured Log Files | 549 |
| <i>Piyush Nimbalkar, Varish Mulwad, Nikhil Puranik, Anupam Joshi, and Tim Finin</i> | |

| | |
|---|-----|
| Semantic Concept Detection Using Weighted Discretization Multiple Correspondence Analysis for Disaster Information Management | 556 |
| <i>Samira Pouyanfar and Shu-Ching Chen</i> | |

Session C12: Computing and Users

| | |
|---|-----|
| A Computational Model for Dyadic Relationships (Invited Paper) | 565 |
| <i>Clio Andris</i> | |
| Reusing the NP-Hard Traveling-Salesman Problem to Demonstrate That P~NP (Invited Paper) | 574 |
| <i>Stuart H. Rubin, Thouraya Bouabana-Tebibel, Yasmin Hoadjli, and Zahira Ghalem</i> | |
| Towards Extracting Coherent User Concerns and Their Hierarchical Organization from User Reviews | 582 |
| <i>Ligaj Pradhan, Chengcui Zhang, and Steven Bethard</i> | |

Session C13: Security and Knowledge Management

| | |
|--|-----|
| ARM: Authenticated Approximate Record Matching for Outsourced Databases | 591 |
| <i>Boxiang Dong and Wendy Wang</i> | |
| Enhancing Multimedia Imbalanced Concept Detection Using VIMP in Random Forests | 601 |
| <i>Saad Sadiq, Yilin Yan, Mei-Ling Shyu, Shu-Ching Chen, and Hemant Ishwaran</i> | |
| Understanding Anti-forensic Techniques with Timestamp Manipulation (Invited Paper) | 609 |
| <i>Dae-il Jang, Gail-Joon Ahn, Hyunuk Hwang, and Kibom Kim</i> | |
| A Classifier Ensemble Framework for Multimedia Big Data Classification | 615 |
| <i>Yilin Yan, Qiusha Zhu, Mei-Ling Shyu, and Shu-Ching Chen</i> | |

IEEE International Workshop on Empirical Methods for Recognizing Inference in Text (EMRITE)

| | |
|---|-----|
| A Real Case Analytics on Social Network of Opinion Spammers | 623 |
| <i>Chih-Chien Wang, Min-Yuh Day, and Yu-Ruei Lin</i> | |
| Mining of Learning Interaction Patterns in MOOCs Forums | 631 |
| <i>K.-C. Chu, C.-S. Wang, and J.-H. Kuo</i> | |
| A Comparison Between Supervised and Unsupervised Models for Identify a Large Number of Categories | 638 |
| <i>Patrícia Andrade, Marcelo Ladeira, and Rommel Carvalho</i> | |

| | |
|---------------------------|-----|
| Author Index | 642 |
|---------------------------|-----|