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

Pittsburgh, Pennsylvania, USA
28-30 July 2016
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

Bhavani Thuraisingham, Murat Kantarcioglu, Kevin Hamlen, Latifur Khan,
Tim Finin, Anupam Joshi, Tim Oates, and Elisa Bertino

Session A11: Security and Privacy
A Novel Method for Fraudulent Medicare Claims Detection from Expected Payment Deviations (Application Paper) .................................................................11

Richard A. Bauder and Taghi M. Khoshgoftaar

Anomaly Detection Techniques for Database Protection Against Insider Threats (Invited Paper) ........................................................................................................20

Asmaa Sallam, Qian Xiao, Elisa Bertino, and Daren Fadolalkarim

Improving the Utility of Anonymized Datasets through Dynamic Evaluation of Generalization Hierarchies .................................................................30

Vanessa Ayala-Rivera, Thomas Cerqueus, Liam Murphy, and Christina Thorpe

A Risk-Benefit Driven Architecture for Personal Data Release (Invited Paper) .................................................................40

B. C. Singh, B. Carminati, and E. Ferrari
Session A12: Reuse and Modeling

Reusable Meta-Models for Crowdsourcing Driven Elastic Systems (Invited Paper) .................................................................50
  Julian Jarrett and M. Brian Blake

Modeling Integration and Reuse of Heterogeneous Terminologies in Faceted Browsing Systems ...................................................58
  Daniel R. Harris

Discover Tipping Users for Cross Network Influencing (Invited Paper) ....................................................................................67
  Qianyi Zhan, Jiawei Zhang, Philip S. Yu, Sherry Emery, and Junyuan Xie

Session A13: Spatial and Mobile Systems

A Machine Learning Approach to Improve the Accuracy of GPS-Based Map-Matching Algorithms (Invited Paper) ........................77
  Mahdi Hashemi and Hassan A. Karimi

An Automatic Approach for Discovering and Geocoding Locations in Domain-Specific Web Data (Application Paper) .......................87
  Chris A. Mattmann and Madhav Sharan

GeoTeCS: Exploiting Geographical, Temporal, Categorical and Social Aspects for Personalized POI Recommendation (Invited Paper) ..............................................................................................94
  Ramesh Baral, Dingding Wang, Tao Li, and Shu-Ching Chen

Predicting Students' Behavioral Patterns in University Networks for Efficient Bandwidth Allocation: A Hybrid Data Mining Method (Application Paper) .........................................................102
  Elham Akhond Zadeh Noughabi, Behrouz H. Far, and Bijan Raahemi

IEEE International Workshop on Formal Methods Integration (FMi)

Modular Verification of Termination and Execution Time Bounds Using Separation Logic ...................................................................110
  Jafar Hamin and Bart Jacobs

Modelling Complex Timing Requirements with Refinement .......................................................................................................118
  Gintautas Sulskus, Michael Poppleton, and Abdolbaghi Rezazadeh

Spekl: A Layered System for Specification Authoring, Sharing, and Usage ..................................................................................126
  John L. Singleton and Gary T. Leavens

A Formal Language for Writing Contracts ...............................................................................................................................134
  William M. Farmer and Qian Hu
Session A21: Classification & Processing

Monitoring Classification Blindspots to Detect Drifts from Unlabeled Data .................................................. 142
   Tegjyot Singh Sethi, Mehmed Kantardzic, and Elaheh Arabmakki

Designing a Testing Framework for Transfer Learning Algorithms (Application Paper) ........................................... 152
   Karl R. Weiss, Taghi M. Khoshgoftaar, and Oneeb Rehman

Cross-Domain Sentiment Analysis: An Empirical Investigation .............................................................................. 160
   Brian Heredia, Taghi M. Khoshgoftaar, Joseph Prusa, and Michael Crawford

Session A22: Learning

Classifying Questions into Fine-Grained Categories Using Topic Enriching .............................................................. 166
   Zhao Lu, Yu-Ru Lin, Qing Zhang, and Mengwei Chen

Clustering Web Pages Based on Structure and Style Similarity (Application Paper) ........................................ 175
   Thamme Gowda and Chris A. Mattmann

Automatic Label Data Abstraction Based on Information Entropy (Application Paper) ........................................ 181
   Fumiya Kudo, Norihiko Moriwaki, Tomoaki Akitomi, Susumu Serita, and Yu Kitano

Algorithmic Decision Support for Personalized Education (Invited Paper) ............................................................. 188
   Tyler Morrow, Sahra Sedigh Sarvestani, and Ali R. Hurson

Session A23: Database and Cloud Systems

Employing Graph Databases as a Standardization Model towards Addressing Heterogeneity ........................................ 198
   Dippy Aggarwal and Karen C. Davis

Query-Driven Sampling for Collective Entity Resolution .............................................................................................. 208
   Christan Grant, Daisy Zhe Wang, and Michael Wick

An Extensible Software Platform for Cloud-Based Decision Support and Automation in Precision Agriculture .................................................. 218
   Li Tan, Hongfei Hou, and Qin Zhang

Cloud-Based Web Services for Endangered Language Analysis (Invited Paper) ......................................................... 226
   Ashik Samad, Kevin Wu, and Min Chen

IEEE International Workshop on Formal Methods Integration (FMi)

Toward Bringing Distributed System Design upon Rigorous Footing ........................................................................ 233
   Mohammed S. Al-Mahfoudh, Ganesh Gopalakrishnan, and Ryan Stutsman
Support for Model Checking Z Specifications .................................................................241
   Maria Ulfah Siregar

BHive: Towards Behaviour-Driven Development Supported by B-Method ..........................249
   J. Carter and W. B. Gardner

Poster Session

Ensemble Classifier for Imbalanced Streaming Data Using Partial Labeling ..................257
   Elaheh Arabmakki, Mehmed Kantardzic, and Tegjyot Singh Sethi

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

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

Analysis of Soccer Robot Behaviors Using Time Petri Nets ........................................270
   Daniel Ponsini, Yilin Yang, and Seung-Yun Kim

Datom Continued: Towards Multi-Objective Optimization of Data Management Entities .................................................................275
   Verena Kantere

Intrusion Detection Using a Multiple-Detector Set Artificial Immune System ..................283
   James Brown, Mohd Anwar, and Gerry Dozier

Session B11: Access Control and Trust

PETRA: Personalized Trust Management Architecture (Application Paper) ....................287
   Hisham Salah and Mohamed Eltoweissy

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

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

General Methods for Access Control Policy Verification (Application Paper) ..................315
   Vincent C. Hu and D. Richard Kuhn
Session B12: Infrastructure Resilience and Disaster Response

A Scalable and Automatic Validation Process for Florida Public Hurricane Loss Model (Invited Paper) ..........................................................324
   Haiman Tian, Hsin-Yu Ha, Samira Pouyanfar, Yilin Yan, Sheng Guan,
   Shu-Ching Chen, Mei-Ling Shyu, and Shahid Hamid

Integrating Control and Fault-Tolerant Wireless Network Design for Small Modular Nuclear Reactors (Invited Paper) ........................................................332
   Wenchen Wang, Christopher D'Angelo, Daniel Mosse, and Daniel Cole

Identification of Power Line Outages Based on PMU Measurements and Sparse Overcomplete Representation ..........................................................343
   Long Cheng, Chenyu You, and Liguo Chen

Selective Mapping for Real-Time Video Transport over Wireless LAN ..........................................................350
   Hongli Luo

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

Generating Headline Candidates for News Articles ..........................................................355
   Naoya Okumura and Takao Miura

Process Discovery Using Classification Tree Hidden Semi-Markov Model ..........................................................361
   Yihuang Kang and Vladimir Zadorozhny

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

Predicting Cancer Relapse with Clinical Data: A Survey of Current Techniques ..........................................................369
   Aaron N. Richter and Taghi M. Khoshgoftaar

Investigating the Variation of Ensemble Size on Bagging-Based Classifier Performance in Imbalanced Bioinformatics Datasets ..........................................................377
   Alireza Fazelpour, Taghi M. Khoshgoftaar, David J. Dittman, and Amri Naplitano

Session B21: Information Processing and Reuse

An Approach to Ontology Integration for Ontology Reuse ..........................................................384
   Enrico G. Caldarola and Antonio M. Rinaldi

Information Fusion Using Characteristic Linear System in Multi-robot Search and Rescue Task ..........................................................394
   Pei-Ju Lee and Vladimir Zadorozhny

Bloom Filter Tree for Fast Search and Synchronization of Tree-Structured Data ..........................................................401
   Mengyu Wang and Ying Zhu
Designing a Better Data Representation for Deep Neural Networks and Text Classification .................................................................411

Joseph D. Prusa and Taghi M. Khoshgoftaar

**Session B22: Data Storage and Analytics**

Data Analytics Using Ontologies of Management Theories: Towards Implementing "From Theory to Practice" (Application Paper) ..........................................................417

*Henry M. Kim, Jackie Ho Nam Cheung, Marek Laskowski, and Iryna Gel*

A Semantically-Enabled SOA Governance Repository (Application Paper) ..................................................................................423

*Jan Königsberger and Bernhard Mitschang*

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

*Yunkai Liu, A. Kate MacPhedran, and Yu Luo*

Domain Knowledge Assisted Data Processing for Florida Public Hurricane Loss Model (Invited Paper) ..................................................................................441

*Yilin Yan, Samira Pouyanfar, Haiman Tian, Sheng Guan, Hsin-Yu Ha, Shu-Ching Chen, Mei-Ling Shyu, and Shahid Hamid*

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

Rapid Prototyping of a Text Mining Application for Cryptocurrency Market Intelligence ........................................................................448

*Marek Laskowski and Henry M. Kim*

Discovering News Frames: Exploring Text, Content, and Concepts in Online News Sources to Address Water Insecurity in the Southwest Region ........................................454

*Loretta H. Cheeks, Tracy L. Stepien, and Dara M. Wald*

A Topic-Independent Hybrid Approach for Sentiment Analysis of Chinese Microblog ........................................................................463

*Ping Han, Shan Li, and Yunfei Jia*

Fast Document Indexing Using Aho-Corasick State Machine .................................................................................................................469

*Ramanpreet Singh and Gurpreet Singh Mohaar*

Recommender System Framework for Academic Choices: Personality Based Recommendation Engine (PBRE) ..................................................476

*Muhammad Fahim Uddin, Soumita Banerjee, and Jeongkyu Lee*
IEEE International Workshop on Information Integration in Cyber Physical Systems (IICPS)

Comparison of Real-Time, Mobile Data Bus Architectures .................................................................484
Paul M. Simon and Scott R. Graham

An Event Detection Framework Supported by a Smart Graphical User Interface ..........................................490
Fabio Persia and Sven Helmer

Tracked Robot over a Slope Path: Dynamic Stability Control ........................................................................496
Daniela D'Auria, Gianluca Ristorto, Gallo Raimondo, and Fabrizio Mazzetto

Anomaly Detection in Electrical Substation Circuits via Unsupervised Machine Learning ........................................500
Alfonso Valdes, Richard Macwan, and Matthew Backes

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
Chi-Ting Liu, Shao-Heng Chen, and Chuan-Jie Lin

A Study on Machine Learning for Imbalanced Datasets with Answer Validation of Question Answering ............513
Min-Yuh Day and Cheng-Chia Tsai

Principle-Based Approach for Semi-Automatic Construction of a Restaurant Question Answering System from Limited Datasets .................................................................520
Ting-Hao Yang, Yu-Lun Hsieh, You-Shan Chung, Cheng-Wei Shih, Shih-Hung Liu, Yung-Chun Chang, and Wen-Lian Hsu

Improving the Selection Error Recognition in a Chinese Grammar Error Detection System .........................................................525
Po-Lin Chen, Shih-Hung Wu, Liang-Pu Chen, and Ping-Che Yang

Session C11: Semantics for Reuse and Integration

Reusing Software Libraries Using Semantic Graphs .....................................................................................531
Awny Alnusair, Majdi Rawashdeh, Mohammed F. Alhamid, M. Anwar Hossain, and Ghulam Muhammad

Knowledge Induction Based on Randomization in Case-Based Reasoning ..................................................541
Thouraya Bouabana-Tebibel, Stuart H. Rubin, Asmaa Chebba, Sofiane Bediar, and Syphax Iskounen

Semantic Interpretation of Structured Log Files .....................................................................................549
Piyush Nimbalkar, Varish Mulwad, Nikhil Puranik, Anupam Joshi, and Tim Finin
Semantic Concept Detection Using Weighted Discretization Multiple Correspondence Analysis for Disaster Information Management .........................................................556
   Samira Pouyanfar and Shu-Ching Chen

**Session C12: Computing and Users**

A Computational Model for Dyadic Relationships (Invited Paper) .................................................................565
   Clio Andris

Reusing the NP-Hard Traveling-Salesman Problem to Demonstrate That P~NP (Invited Paper) ........................................................................................................574
   Stuart H. Rubin, Thouraya Bouabana-Tebibel, Yasmin Hoadjli, and Zahira Ghalem

Towards Extracting Coherent User Concerns and Their Hierarchical Organization from User Reviews .................................................................582
   Ligaj Pradhan, Chengcui Zhang, and Steven Bethard

**Session C13: Security and Knowledge Management**

ARM: Authenticated Approximate Record Matching for Outsourced Databases ..................................................591
   Boxiang Dong and Wendy Wang

Enhancing Multimedia Imbalanced Concept Detection Using VIMP in Random Forests ........................................601
   Saad Sadiq, Yilin Yan, Mei-Ling Shyu, Shu-Ching Chen, and Hemant Ishwaran

Understanding Anti-forensic Techniques with Timestamp Manipulation (Invited Paper) ........................................609
   Dae-il Jang, Gail-Joon Ahn, Hyunuk Hwang, and Kibom Kim

A Classifier Ensemble Framework for Multimedia Big Data Classification ....................................................615
   Yilin Yan, Qiusha Zhu, Mei-Ling Shyu, and Shu-Ching Chen


A Real Case Analytics on Social Network of Opinion Spammers .................................................................623
   Chih-Chien Wang, Min-Yuh Day, and Yu-Ruei Lin

Mining of Learning Interaction Patterns in MOOCs Forums ........................................................................631
   K.-C. Chu, C.-S. Wang, and J.-H. Kuo

A Comparison Between Supervised and Unsupervised Models for Identify a Large Number of Categories ........................................................................638
   Patricia Andrade, Marcelo Ladeira, and Rommel Carvalho

**Author Index** ............................................................................................................................................642