# Contents

## Preface
Message from the Chairs ................................................................. iii

## Keynote
What Is Software Development Productivity, Anyway? (Keynote)  
Gail C. Murphy — University of British Columbia, Canada  ................................................................. 1

## Bug Triaging
Why So Complicated? Simple Term Filtering and Weighting for Location-Based Bug Report Assignment Recommendation  
Ramin Shokripour, John Anvik, Zarinah M. Kasirun, and Sima Zamani — University of Malaya, Malaysia; Central Washington University, USA  ................................................................. 2

Which Work-Item Updates Need Your Response?  
Debdoot Mukherjee and Malika Garg — IBM Research, India; IIT Delhi, India  ................................................................. 12

Bug Report Assignee Recommendation using Activity Profiles  
Hoda Naguib, Nitesh Narayan, Bernd Brügge, and Dina Helal — TU Munich, Germany; German University in Cairo, Egypt  ................................................................. 22

## MSR Goes Mobile
Asking for (and about) Permissions Used by Android Apps  
Ryan Stevens, Jonathan Ganz, Vladimir Filkov, Premkumar Devanbu, and Hao Chen — UC Davis, USA  ................................................................. 31

Retrieving and Analyzing Mobile Apps Feature Requests from Online Reviews  
Claudia Iacob and Rachel Harrison — Oxford Brookes University, UK  ................................................................. 41

Gerrit Software Code Review Data from Android  
Murtuza Mukadam, Christian Bird, and Peter C. Rigby — Concordia University, Canada; Microsoft Research, USA  ................................................................. 45

Who Does What during a Code Review? Datasets of OSS Peer Review Repositories  
Kazuki Hamasaki, Raula Gaikovina Kula, Norihiro Yoshida, A. E. Camargo Cruz, Kenji Fujiwara, and Hajimu Iida — NAIST, Japan; Osaka University, Japan  ................................................................. 49

## MSR Challenge
Why, When, and What: Analyzing Stack Overflow Questions by Topic, Type, and Code  
Miltiadis Allamanis and Charles Sutton — University of Edinburgh, UK  ................................................................. 53

Deficient Documentation Detection: A Methodology to Locate Deficient Project Documentation using Topic Analysis  
Joshua Charles Campbell, Chenlei Zhang, Zhen Xu, Abram Hindle, and James Miller — University of Alberta, Canada  ................................................................. 57

Detecting API Usage Obstacles: A Study of iOS and Android Developer Questions  
Wei Wang and Michael W. Godfrey — University of Waterloo, Canada  ................................................................. 61

Encouraging User Behaviour with Achievements: An Empirical Study  
Scott Grant and Buddy Betts — Queen’s University, Canada; OUYA, USA  ................................................................. 65

Is Programming Knowledge Related to Age? An Exploration of Stack Overflow  
Patrick Morrison and Emerson Murphy-Hill — North Carolina State University, USA  ................................................................. 69
A Discriminative Model Approach for Suggesting Tags Automatically for Stack Overflow Questions
Avigit K. Saha, Ripon K. Saha, and Kevin A. Schneider — University of Saskatchewan, Canada; University of Texas at Austin, USA

Exploring Activeness of Users in QA Forums
Vibha Singhal Sinha, Senthil Mani, and Monika Gupta — IBM Research, India

A Study of Innovation Diffusion through Link Sharing on Stack Overflow
Carlos Gómez, Brendan Cleary, and Leif Singer — University of Victoria, Canada

Making Sense of Online Code Snippets
Siddharth Subramanian and Reid Holmes — University of Waterloo, Canada

Building Reputation in StackOverflow: An Empirical Investigation
Amiangshu Bosu, Christopher S. Corley, Dustin Heaton, Debarshi Chatterji, Jeffrey C. Carver, and Nicholas A. Kraft — University of Alabama, USA

An Exploratory Analysis of Mobile Development Issues using Stack Overflow
Mario Linares-Vásquez, Bogdan Dit, and Denys Poshyvanyk — College of William and Mary, USA

Answering Questions about Unanswered Questions of Stack Overflow
Muhammad Asaduzzaman, Ahmed Shah Mashiyat, Chanchal K. Roy, and Kevin A. Schneider — University of Saskatchewan, Canada; University of Toronto, Canada

Changes and Fixes
Will My Patch Make It? And How Fast?: Case Study on the Linux Kernel
Yujuan Jiang, Bram Adams, and Daniel M. German — Polytechnique Montréal, Canada; University of Victoria, Canada

Linux Variability Anomalies: What Causes Them and How Do They Get Fixed?
Sarah Nadi, Christian Dietrich, Reinhard Tartler, Richard C. Holt, and Daniel Lohmann — University of Waterloo, Canada; University of Erlangen-Nuremberg, Germany

The Impact of Tangled Code Changes
Kim Herzig and Andreas Zeller — Microsoft Research, UK; Saarland University, Germany

A Dataset from Change History to Support Evaluation of Software Maintenance Tasks
Bogdan Dit, Andrew Holtzhauer, Denys Poshyvanyk, and Huzefa Kagdi — College of William and Mary, USA; Wichita State University, USA

Apache Commits: Social Network Dataset
Alexander C. MacLean and Charles D. Knutson — Brigham Young University, USA

Software Evolution
Understanding the Evolution of Type-3 Clones: An Exploratory Study
Ripon K. Saha, Chanchal K. Roy, Kevin A. Schneider, and Dewayne E. Perry — University of Texas at Austin, USA; University of Saskatchewan, Canada

An Empirical Study of the Fault-Proneness of Clone Mutation and Clone Migration
Shuai Xie, Foutse Khomh, and Ying Zou — Queen’s University, Canada; Polytechnique Montréal, Canada

Intensive Metrics for the Study of the Evolution of Open Source Projects: Case Studies from Apache Software Foundation Projects
Santiago Gala-Pérez, Gregorio Robles, Jesús M. González-Barahona, and Israel Herraiz — Apache Software Foundation, Spain; Universidad Rey Juan Carlos, Spain; Universidad Politécnica de Madrid, Spain

A Preliminary Investigation of Using Age and Distance Measures in the Detection of Evolutionary Couplings
Abdulkareem Alali, Brian Bartman, Christian D. Newman, and Jonathan I. Maletic — Kent State University, USA

Analysis of Bug Reports
Search-Based Duplicate Defect Detection: An Industrial Experience
Mehdi Amoui, Nilam Kaushik, Abraham Al-Dabbagh, Ladan Tahvildari, Shimin Li, and Weining Liu — University of Waterloo, Canada; BlackBerry, Canada

A Contextual Approach towards More Accurate Duplicate Bug Report Detection
Anahita Alipour, Abram Hindle, and Eleni Stroulia — University of Alberta, Canada
Bug Resolution Catalysts: Identifying Essential Non-committers from Bug Repositories
Senthil Mani, Seema Nagar, Debdeep Mukherjee, Ramasuri Narayanan, Vibha Singhal Sinha, and Amit A. Nanavati — IBM Research, India

The Eclipse and Mozilla Defect Tracking Dataset: A Genuine Dataset for Mining Bug Information
Ahmed Lamkanfi, Javier Pérez, and Serge Demeyer — University of Antwerp, Belgium

Software Ecosystems, Big Data
Mining Source Code Repositories at Massive Scale using Language Modeling
Miltiadis Allamanis and Charles Sutton — University of Edinburgh, UK

Do Software Categories Impact Coupling Metrics?
Lucas Batista Leite de Souza and Marcelo de Almeida Maia — UFU, Brazil

The Maven Repository Dataset of Metrics, Changes, and Dependencies
Steven Raemaekers, Arie van Deursen, and Joost Visser — Software Improvement Group, Netherlands; TU Delft, Netherlands

A Historical Dataset for the Gnome Ecosystem
Mathieu Goeminne, Maëlick Claes, and Tom Mens — University of Mons, Belgium

A Network of Rails: A Graph Dataset of Ruby on Rails and Associated Projects
Patrick Wagstrom, Corey Jergensen, and Anita Sarma — IBM Research, USA; University of Nebraska-Lincoln, USA

The GHTorrent Dataset and Tool Suite
Georgios Gousios — TU Delft, Netherlands

Bug/Change Classification and Localization
Discovering, Reporting, and Fixing Performance Bugs
Adrian Nistor, Tian Jiang, and Lin Tan — University of Illinois at Urbana-Champaign, USA; University of Waterloo, Canada

Improving Bug Localization using Correlations in Crash Reports
Shaohua Wang, Foutse Khomh, and Ying Zou — Queen’s University, Canada; Polytechnique Montréal, Canada

Steven Raemaekers, Gabriela F. Nane, Arie van Deursen, and Joost Visser — Software Improvement Group, Netherlands; TU Delft, Netherlands

Social Mining
Fixing the ‘Out of Sight Out of Mind’ Problem: One Year of Mood-Based Microblogging in a Distributed Software Team
Kevin Dullemond, Ben van Gameren, Margaret-Anne Storey, and Arie van Deursen — TU Delft, Netherlands; University of Victoria, Canada

Communication in Open Source Software Development Mailing Lists
Anja Guzzi, Alberto Bacchelli, Michele Lanza, Martin Pinzger, and Arie van Deursen — TU Delft, Netherlands; University of Lugano, Switzerland; University of Klagenfurt, Austria

Tag Recommendation in Software Information Sites
Xin Xia, David Lo, Xinyu Wang, and Bo Zhou — Zhejiang University, China; Singapore Management University, Singapore

Using Developer Interaction Data to Compare Expertise Metrics
Romain Robbes and David Röthlisberger — University of Chile, Chile; Federico Santa María Technical University, Chile

Project Roles in the Apache Software Foundation: A Dataset
Megan Squire — Elon University, USA

Apache-Affiliated Twitter Screen Names: A Dataset
Megan Squire — Elon University, USA

Search-Driven Development
Assisting Code Search with Automatic Query Reformulation for Bug Localization
Bunyamin Sisman and Avinash C. Kak — Purdue University, USA

Mining Succinct and High-Coverage API Usage Patterns from Source Code
Jue Wang, Yingnong Dang, Hongyu Zhang, Kai Chen, Tao Xie, and Dongmei Zhang — Tsinghua University, China; Microsoft Research, China; Peking University, China; North Carolina State University, USA
Rendezvous: A Search Engine for Binary Code
Wei Ming Khoo, Alan Mycroft, and Ross Anderson — University of Cambridge, UK .................................................. 329

An Unabridged Source Code Dataset for Research in Software Reuse
Werner Janjic, Oliver Hummel, Marcus Schumacher, and Colin Atkinson — University of Mannheim, Germany; KIT, Germany 339

10 Years of MSR
The MSR Cookbook: Mining a Decade of Research
Hadi Hemmati, Sarah Nadi, Olga Baysal, Oleksii Kononenko, Wei Wang, Reid Holmes, and Michael W. Godfrey — University of Waterloo, Canada .......................................................... 343

Happy Birthday! A Trend Analysis on Past MSR Papers
Serge Demeyer, Alessandro Murgia, Kevin Wyckmans, and Ahmed Lamkanfi — University of Antwerp, Belgium ........ 353

Replicating Mining Studies with SOFAS
Giacomo Ghezzi and Harald C. Gall — University of Zurich, Switzerland ................................................................. 363

A Historical Dataset of Software Engineering Conferences
Bogdan Vasilescu, Alexander Serebrenik, and Tom Mens — TU Eindhoven, Netherlands; University of Mons, Belgium .... 373

Mining Unstructured Data
Automatically Mining Software-Based, Semantically-Similar Words from Comment-Code Mappings
Matthew J. Howard, Samir Gupta, Lori Pollock, and K. Vijay-Shanker — University of Delaware, USA ....................... 377

Strategies for Avoiding Text Fixture Smells during Software Evolution
Michaela Greiler, Andy Zaidman, Arie van Deursen, and Margaret-Anne Storey — TU Delft, Netherlands; University of Victoria, Canada ................................................................. 387

Contextual Analysis of Program Logs for Understanding System Behaviors
Qiang Fu, Jian-Guang Lou, Qingwei Lin, Rui Ding, Dongmei Zhang, and Tao Xie — Microsoft Research, China; Microsoft, China; North Carolina State University, USA ................................. 397

A Dataset for Evaluating Identifier Splitters
David Binkley, Dawn Lawrie, Lori Pollock, Emily Hill, and K. Vijay-Shanker — Loyola University Maryland, USA; University of Delaware, USA; Montclair State University, USA ................................................. 401

INVocD: Identifier Name Vocabulary Dataset
Simon Butler, Michel Wermelinger, Yijun Yu, and Helen Sharp — Open University, UK .................................................. 405

Predictor Models
Better Cross Company Defect Prediction
Fayola Peters, Tim Menzies, and Andrian Marcus — West Virginia University, USA; Wayne State University, USA .... 409

Using Citation Influence to Predict Software Defects
Wei Hu and Kenny Wong — University of Alberta, Canada .............................................................................. 419

Revisiting Software Development Effort Estimation Based on Early Phase Development Activities
Masateru Tsunoda, Koji Toda, Kyohei Fushida, Yasutaka Kamei, Meiyappan Nagappan, and Naoyasu Uabayashi — Toyo University, Japan; Fukuoka Institute of Technology, Japan; NTT, Japan; Kyushu University, Japan; Queen’s University, Canada . 429

Author Index