2017 IEEE/ACM 14th International Conference on Mining Software Repositories (MSR 2017)

Buenos Aires, Argentina
20 – 21 May 2017
Table of Contents

Message from MSR 2017 General Chairs ........................................................................................................... xii
MSR 2017 Organizing Committee .................................................................................................................... xv
MSR 2017 Program Committee ........................................................................................................................ xvi
MSR 2017 Data Showcase Program Committee ............................................................................................ xviii
MSR 2017 Mining Challenge Program Committee ......................................................................................... xix
MSR 2017 Steering Committee ......................................................................................................................... xx
MSR 2017 Reviewers ......................................................................................................................................... xxi
ICSE 2017 Sponsors and Benefactors ............................................................................................................... xxiii

Keynote

Half-Century of Unix: History, Preservation, and Lessons Learned ................................................................. 1
   Diomidis Spinellis
   — Athens University of Economics and Business

Mobile

An Empirical Study on Android-Related Vulnerabilities .................................................................................... 2
   Mario Linares-Vásquez, Gabriele Bavota, and Camilo Escobar-Velásquez
   — Universidad de los Andes; Università della Svizzera Italiana

Understanding the Origins of Mobile App Vulnerabilities: A Large-Scale Measurement
Study of Free and Paid Apps ................................................................................................................................ 14
   Takuya Watanabe, Mitsuaki Akiyama, Fumihiro Kanei, Eitaro Shioji, Yuta Takata,
   Bo Sun, Yuta Ishi, Toshiki Shibahara, Takeshi Yagi, and Tatsuya Mori
   — NTT Secure Platform Laboratories; Waseda University

Developer Mistakes in Writing Android Manifests: An Empirical Study
of Configuration Errors ........................................................................................................................................ 25
   Ajay Kumar Jha, Sunghee Lee, and Woo Jin Lee
   — Kyungpook National University

How Do Apps Evolve in Their Permission Requests? A Preliminary Study ...................................................... 37
   Paolo Calciati and Alessandra Gorla
   — IMDEA Software Institute
A Study on the Energy Consumption of Android App Development Approaches ........................................ 42
Wellington Oliveira, Renato Oliveira, and Fernando Castor
— Federal University of Pernambuco

Candoia: A Platform for Building and Sharing Mining Software Repositories Tools as Apps .................... 53
Nitin M. Tiwari, Ganesha Upadhyaya, Hoan Anh Nguyen, and Hridesh Rajan
— Iowa State University

Dependencies

Analyzing Program Dependencies in Java EE Applications........................................................................ 64
Anas Shatnawi, Hafedh Mili, Ghizlane El Boussaidi, Anis Boubaker,
Yann-Gaël Guéhéneuc, Naouel Moha, Jean Privat,
and Manel Abdellatif
— Université du Québec à Montréal

Mining Social Web Service Repositories for Social Relationships
to Aid Service Discovery ......................................................................................................................... 75
Alejandro Corbellini, Daniela Godoy, Cristian Mateos, Alejandro Zunino,
and Ignacio Lizarralde
— ISISTAN-CONICET, UNICEN

Who You Gonna Call? Analyzing Web Requests in Android Applications ................................................. 80
Marianna Rapoport, Philippe Suter, Erik Wittern, Ondrej Lhotak,
and Julian Dolby
— University of Waterloo; IBM T.J. Watson Research Center

Extracting Code Segments and Their Descriptions from Research Articles .............................................. 91
Preetha Chatterjee, Benjamin Gause, Hunter Hedinger, and Lori Pollock
— University of Delaware

Structure and Evolution of Package Dependency Networks ...................................................................... 102
Riivo Kikas, Georgios Gousios, Marlon Dumas, and Dietmar Pfahl
— University of Tartu; Delft University of Technology

Spencer: Interactive Heap Analysis for the Masses ................................................................................. 113
Stephan Brandauer and Tobias Wrigstad
— Uppsala University

Modelling and Prediction

Predicting Likelihood of Requirement Implementation within the Planned Iteration:
An Empirical Study at IBM ........................................................................................................................ 124
Ali Dehghan, Adam Neal, Kelly Blincoe, Johan Linaker, and Daniela Damian
— University of Victoria; Persistent Systems; University of Auckland;
Lund University

The Impact of Using Regression Models to Build Defect Classifiers..................................................... 135
Gopi Krishnan Rajbahadur, Shaowei Wang, Yasutaka Kamei,
and Ahmed E. Hassan
— Queen’s University; Kyushu University
A Large-Scale Study of the Impact of Feature Selection Techniques on Defect Classification Models ......................................................................................................................................................................................... 146

Baljinder Ghotra, Shane McIntosh, and Ahmed E. Hassan
— Queen’s University

SpreadCluster: Recovering Versioned Spreadsheets through Similarity-Based Clustering ................ 158
Liang Xu, Wensheng Dou, Chushu Gao, Jie Wang, Jun Wei, Hua Zhong, and Tao Huang
— Chinese Academy of Sciences

Who Will Leave the Company?: A Large-Scale Industry Study of Developer Turnover by Mining Monthly Work Report ......................................................................................................................................................................................... 170
Lingfeng Bao, Zhencang Xing, Xin Xia, David Lo, and Shanping Li
— Zhejiang University; Australian National University; University of British Columbia; Singapore Management University

Concept-Based Classification of Software Defect Reports ................................................................ 182
Sangameshwar Patil
— Tata Research Development and Design Centre

NLP and Code Review

Choosing an NLP Library for Analyzing Software Documentation: A Systematic Literature Review and a Series of Experiments ......................................................................................................................................................................................... 187
Fouad Nasser A Al Omran and Christoph Treude
— University of Adelaide

Bootstrapping a Lexicon for Emotional Arousal in Software Engineering ............................................. 198
Mika V. Mäntylä, Nicole Novielli, Filippo Lanubile, Maëlick Claes, and Miikka Kuutila
— University of Oulu; University of Bari

Leveraging Automated Sentiment Analysis in Software Engineering ..................................................... 203
Md Rakibul Islam and Minhaz F. Zibran
— University of New Orleans

Predicting Usefulness of Code Review Comments Using Textual Features and Developer Experience ......................................................................................................................................................................................................................................................... 215
Mohammad Masudur Rahman, Chanchal K. Roy, and Raula G. Kula
— University of Saskatchewan; Osaka University

Classifying Code Comments in Java Open-Source Software Systems ..................................................... 227
Luca Pascarella and Alberto Bacchelli
— Delft University of Technology

Using Q&A Websites as a Method for Assessing Systematic Reviews ..................................................... 238
Bruno Cartaxo, Gustavo Pinto, Danilo Ribeiro, Fernando Kamei, Ronnie E.S. Santos, Fábio Q.B. Da Silva, and Sérgio Soares
— Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco; Universidade Federal de Pernambuco; Instituto Federal do Pará; Instituto Federal de Alagoas
Abnormal Working Hours: Effect of Rapid Releases and Implications to Work Content .......................... 243
Maëlick Claes, Mika Mäntylä, Miikka Kuutila, and Bram Adams
— University of Oulu; École Polytechnique de Montréal

Clones and Edits

Mining Change Histories for Unknown Systematic Edits .......................................................................... 248
Tim Molderez, Reinout Stevens, and Coen De Roover
— Vrije Universiteit Brussel

Source File Set Search for Clone-and-Own Reuse Analysis........................................................................ 257
Takashi Ishio, Yusuke Sakaguchi, Kaoru Ito, and Katsuro Inoue
— Osaka University

RefDiff: Detecting Refactorings in Version Histories ................................................................................ 269
Danilo Silva and Marco Tulio Valente
— Universidade Federal de Minas Gerais

Stack Overflow in Github: Any Snippets There? ....................................................................................... 280
Di Yang, Pedro Martins, Vaibhav Saini, and Cristina Lopes
— University of California, Irvine

Some from Here, Some from There: Cross-Project Code Reuse in GitHub ............................................. 291
Mohammad Gharehyazie, Baishakhi Ray, and Vladimir Filkov
— University of California, Davis; University of Virginia

Exception Evolution in Long-Lived Java Systems .................................................................................... 302
Haidar Osman, Andrei Chis, Claudio Corrodi, Mohammad Ghafari, and Oscar Nierstrasz
— University of Bern; Feenk GmbH

Continuous Integration and Build

Do Not Trust Build Results at Face Value—An Empirical Study of 30 Million CPAN Builds .................... 312
Mahdis Zolfagharinia, Bram Adams, and Yann-Gaël Guéhénué
— École Polytechnique de Montréal

An Empirical Analysis of the Docker Container Ecosystem on GitHub .................................................... 323
Jürgen Cito, Gerald Schermann, John Erik Wittern, Philipp Leitner, Sali Zumberi, and Harald C. Gall
— University of Zurich; IBM T.J. Watson Research Center

How Open Source Projects Use Static Code Analysis Tools in Continuous Integration Pipelines ............ 334
Fiorella Zampetti, Simone Scalabrino, Rocco Oliveto, Gerardo Canfora, and Massmiliano Di Penta
— University of Sannio; University of Molise

An Empirical Analysis of Build Failures in the Continuous Integration Workflows of Java-Based Open-Source Software ................................................................. 345
Thomas Rausch, Waldemar Hummer, Philipp Leitner, and Stefan Schulte
— Vienna University of Technology; University of Zurich
Oops, My Tests Broke the Build: An Explorative Analysis of Travis CI with GitHub
Moritz Beller, Georgios Gousios, and Andy Zaidman
— Delft University of Technology

Extracting Build Changes with BUILDDIFF
Christian Macho, Shane McIntosh, and Martin Pinzger
— University of Klagenfurt; McGill University

Testing and Bugs

An Exploratory Study on Assessing the Impact of Environment Variations on the Results of Load Tests
Ruoyu Gao and Zhen Ming (Jack) Jiang
— York University, Toronto

A Large-Scale Study on the Usage of Testing Patterns That Address Maintainability Attributes: Patterns for Ease of Modification, Diagnoses, and Comprehension
Danielle Gonzalez, Joanna C.S. Santos, Andrew Popovich, Mehdi Mirakhorli, and Mei Nagappan
— Rochester Institute of Technology; University of Waterloo

To Mock or Not to Mock? An Empirical Study on Mocking Practices
Davide Spadini, Mauricio Aniche, Magiel Bruntink, and Alberto Bacchelli
— Software Improvement Group; Delft University of Technology

Bug Characteristics in Blockchain Systems: A Large-Scale Empirical Study
Zhiyuan Wan, David Lo, Xin Xia, and Liang Cai
— Zhejiang University; Singapore Management University; University of British Columbia

Euphony: Harmonious Unification of Cacophonous Anti-Virus Vendor Labels for Android Malware
Médéric Hurier, Guillermo Suarez-Tangil, Santanu Kumar Dash, Tegawendé F. Bissyandé, Yves Le Traon, Jacques Klein, and Lorenzo Cavallaro
— University of Luxembourg; Royal Holloway, University of London

Rationale in Development Chat Messages: An Exploratory Study
Rana Alkadhi, Teodora Lata, Emitza Guzmany, and Bernd Bruegge
— Technische Universität München; University of Zurich

Mining Challenge

TravisTorrent: Synthesizing Travis CI and GitHub for Full-Stack Research on Continuous Integration
Moritz Beller, Georgios Gousios, and Andy Zaidman
— Delft University of Technology

On the Differences between Unit and Integration Testing in the TravisTorrent Dataset
Gerardo Orellana, Gulsher Laghari, Alessandro Murgia, and Serge Demeyer
— University of Antwerp
Cost-Effective Build Outcome Prediction Using Cascaded Classifiers ..................................................... 455
   Ansong Ni and Ming Li
   — Nanjing University

Sentiment Analysis of Travis CI Builds ..................................................................................................... 459
   Rodrigo Souza and Bruno Silva
   — Salvador University; Federal University of Bahia

A Time Series Analysis of TravisTorrent Builds: To Everything There Is a Season .................................... 463
   Abigail Atchison, Christina Berardi, Natalie Best, Elizabeth Stevens,
   and Erik Linstead
   — Chapman University

Insights into Continuous Integration Build Failures ................................................................................... 467
   Md Rakibul Islam and Minhaz F. Zibran
   — University of New Orleans

An Empirical Study of the Personnel Overhead of Continuous Integration .............................................. 471
   Marco Manglaviti, Eduardo Coronado-Montoya, Keheliya Gallaba,
   and Shane McIntosh
   — McGill University

How Does Contributors’ Involvement Influence the Build Status of an Open-Source
Software Project? ....................................................................................................................................... 475
   Marcel Rebouças, Renato O. Santos, Gustavo Pinto, and Fernando Castor
   — Federal University of Pernambuco; Federal Institute of Pará

On the Interplay between Non-Functional Requirements and Builds on Continuous Integration ............... 479
   Klérisson V. R. Paixão, Crícia Z. Felício, Fernanda M. Delfim,
   and Marcelo De A. Maia
   — Universidade Federal de Uberlândia; Instituto Federal do Triângulo Mineiro

Analyzing the Impact of Social Attributes on Commit Integration Success .............................................. 483
   Mauricio Soto, Zack Coker, and Claire Le Goues
   — Carnegie Mellon University

Built to Last or Built Too Fast? Evaluating Prediction Models for Build Times ....................................... 487
   Ekaba Bisong, Eric Tran, and Olga Baysal
   — Carleton University

The Impact of the Adoption of Continuous Integration on Developer Attraction and Retention....................... 491
   Yash Gupta, Yusaira Khan, Keheliya Gallaba, and Shane McIntosh
   — McGill University

An Empirical Study of Activity, Popularity, Size, Testing, and Stability in Continuous Integration ............... 495
   Aakash Gautam, Saket Vishwasrao, and Francisco Servant
   — Virginia Tech

Impact of Continuous Integration on Code Reviews .................................................................................... 499
   Mohammad Masudur Rahman, and Chanchal K. Roy
   — University of Saskatchewan
Prevalence of Botched Code Integrations ................................................................. 503
Ward Muylaert and Coen De Roover
— Vrije Universiteit Brussel

Data Showcase

Software Evolution and Quality Data from Controlled, Multiple, Industrial Case Studies .............. 507
Aiko Yamashita, S. Amirhossein Abtahizadeh, Foutse Khomh,
and Yann-Gaël Guéhéneuc
— Centrum Wiskunde & Informatica; École Polytechnique de Montréal

A Dataset of Scratch Programs: Scraped, Shaped and Scored .................................................... 511
Efthimia Aivaloglou, Felienne Hermans, Jesus Moreno-Leon,
and Gregorio Robles
— Delft University of Technology; Programamos.es;
Universidad Rey Juan Carlos

Continuous Defect Prediction: The Idea and a Related Dataset .................................................. 515
Lech Madeyski and Marcin Kawalerowicz
— Wroclaw University of Science and Technology; Opole University of Technology

An Extensive Dataset of UML Models in GitHub ........................................................................ 519
Gregorio Robles, Truong Ho-Quang, Regina Hebig, Michel R.V. Chaudron,
and Miguel Angel Fernandez
— Universidad Rey Juan Carlos; Chalmers & Gothenburg University

A Dataset for Dynamic Discovery of Semantic Changes in Version Controlled
Software Histories .................................................................................................................. 523
Chenguang Zhu, Yi Li, Julia Rubin, and Marsha Chechik
— University of Toronto; University of British Columbia

Rediscovery Datasets: Connecting Duplicate Reports ................................................................. 527
Mefta Sadat, Ayse Basar Bener, and Andriy Miranskyy
— Ryerson University

A Data Set of OCL Expressions on GitHub .................................................................................. 531
Jeroen Noten, Josh G.M. Mengerink, and Alexander Serebrenik
— Eindhoven University of Technology

Author Index ................................................................................................................................. 535