

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

**Buenos Aires, Argentina  
20 – 21 May 2017**



**IEEE Catalog Number: CFP1778C-POD  
ISBN: 978-1-5386-1545-4**

**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:	CFP1778C-POD
ISBN (Print-On-Demand):	978-1-5386-1545-4
ISBN (Online):	978-1-5386-1544-7

**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](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

CURRAN ASSOCIATES INC.  
**proceedings**  
.com

# 2017 IEEE/ACM 14th International Conference on Mining Software Repositories (MSR 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
<i>Diomidis Spinellis</i> — <i>Athens University of Economics and Business</i>	

### Mobile

An Empirical Study on Android-Related Vulnerabilities .....	2
<i>Mario Linares-Vásquez, Gabriele Bavota, and Camilo Escobar-Velásquez</i> — <i>Universidad de los Andes; Università della Svizzera Italiana</i>	
Understanding the Origins of Mobile App Vulnerabilities: A Large-Scale Measurement Study of Free and Paid Apps .....	14
<i>Takuya Watanabe, Mitsuaki Akiyama, Fumihiko Kanei, Eitaro Shioji, Yuta Takata, Bo Sun, Yuta Ishi, Toshiki Shibahara, Takeshi Yagi, and Tatsuya Mori</i> — <i>NTT Secure Platform Laboratories; Waseda University</i>	
Developer Mistakes in Writing Android Manifests: An Empirical Study of Configuration Errors .....	25
<i>Ajay Kumar Jha, Sunghee Lee, and Woo Jin Lee</i> — <i>Kyungpook National University</i>	
How Do Apps Evolve in Their Permission Requests? A Preliminary Study .....	37
<i>Paolo Calciati and Alessandra Gorla</i> — <i>IMDEA Software Institute</i>	

A Study on the Energy Consumption of Android App Development Approaches .....	42
<i>Wellington Oliveira, Renato Oliveira, and Fernando Castor</i>	
— <i>Federal University of Pernambuco</i>	
Candoia: A Platform for Building and Sharing Mining Software Repositories Tools as Apps .....	53
<i>Nitin M. Tiwari, Ganesha Upadhyaya, Hoan Anh Nguyen, and Hridayesh Rajan</i>	
— <i>Iowa State University</i>	

## Dependencies

Analyzing Program Dependencies in Java EE Applications .....	64
<i>Anas Shatnawi, Hamed Mili, Ghizlane El Boussaidi, Anis Boubaker,</i>	
<i>Yann-Gaël Guéhéneuc, Naouel Moha, Jean Privat,</i>	
<i>and Manel Abdellatif</i>	
— <i>Université du Québec à Montréal</i>	
Mining Social Web Service Repositories for Social Relationships to Aid Service Discovery .....	75
<i>Alejandro Corbellini, Daniela Godoy, Cristian Mateos, Alejandro Zunino,</i>	
<i>and Ignacio Lizarralde</i>	
— <i>ISISTAN-CONICET, UNICEN</i>	
Who You Gonna Call? Analyzing Web Requests in Android Applications .....	80
<i>Marianna Rapoport, Philippe Suter, Erik Wittern, Ondrej Lhotak,</i>	
<i>and Julian Dolby</i>	
— <i>University of Waterloo; IBM T.J. Watson Research Center</i>	
Extracting Code Segments and Their Descriptions from Research Articles .....	91
<i>Preetha Chatterjee, Benjamin Gause, Hunter Hedinger, and Lori Pollock</i>	
— <i>University of Delaware</i>	
Structure and Evolution of Package Dependency Networks .....	102
<i>Riivo Kikas, Georgios Gousios, Marlon Dumas, and Dietmar Pfahl</i>	
— <i>University of Tartu; Delft University of Technology</i>	
Spencer: Interactive Heap Analysis for the Masses .....	113
<i>Stephan Brandauer and Tobias Wrigstad</i>	
— <i>Uppsala University</i>	

## Modelling and Prediction

Predicting Likelihood of Requirement Implementation within the Planned Iteration: An Empirical Study at IBM .....	124
<i>Ali Dehghan, Adam Neal, Kelly Blincoe, Johan Linaker, and Daniela Damian</i>	
— <i>University of Victoria; Persistent Systems; University of Auckland;</i>	
<i>Lund University</i>	
The Impact of Using Regression Models to Build Defect Classifiers .....	135
<i>Gopi Krishnan Rajbahadur, Shaowei Wang, Yasutaka Kamei,</i>	
<i>and Ahmed E. Hassan</i>	
— <i>Queen's University; Kyushu University</i>	

A Large-Scale Study of the Impact of Feature Selection Techniques on Defect Classification Models .....	146
<i>Baljinder Ghotra, Shane McIntosh, and Ahmed E. Hassan</i>	
— <i>Queen's University</i>	
SpreadCluster: Recovering Versioned Spreadsheets through Similarity-Based Clustering .....	158
<i>Liang Xu, Wensheng Dou, Chushu Gao, Jie Wang, Jun Wei, Hua Zhong,</i> <i>and Tao Huang</i>	
— <i>Chinese Academy of Sciences</i>	
Who Will Leave the Company?: A Large-Scale Industry Study of Developer Turnover by Mining Monthly Work Report .....	170
<i>Lingfeng Bao, Zhenchang Xing, Xin Xia, David Lo, and Shanping Li</i>	
— <i>Zhejiang University; Australian National University; University of British Columbia;</i> <i>Singapore Management University</i>	
Concept-Based Classification of Software Defect Reports .....	182
<i>Sangameshwar Patil</i>	
— <i>Tata Research Development and Design Centre</i>	
<b>NLP and Code Review</b>	
Choosing an NLP Library for Analyzing Software Documentation: A Systematic Literature Review and a Series of Experiments .....	187
<i>Fouad Nasser A Al Omran and Christoph Treude</i>	
— <i>University of Adelaide</i>	
Bootstrapping a Lexicon for Emotional Arousal in Software Engineering .....	198
<i>Mika V. Mäntylä, Nicole Novielli, Filippo Lanubile, Maëlick Claes,</i> <i>and Miikka Kuutila</i>	
— <i>University of Oulu; University of Bari</i>	
Leveraging Automated Sentiment Analysis in Software Engineering .....	203
<i>Md Rakibul Islam and Minhaz F. Zibran</i>	
— <i>University of New Orleans</i>	
Predicting Usefulness of Code Review Comments Using Textual Features and Developer Experience .....	215
<i>Mohammad Masudur Rahman, Chanchal K. Roy, and Raula G. Kula</i>	
— <i>University of Saskatchewan; Osaka University</i>	
Classifying Code Comments in Java Open-Source Software Systems .....	227
<i>Luca Pascarella and Alberto Bacchelli</i>	
— <i>Delft University of Technology</i>	
Using Q&A Websites as a Method for Assessing Systematic Reviews .....	238
<i>Bruno Cartaxo, Gustavo Pinto, Danilo Ribeiro, Fernando Kamei, Ronnie E.S. Santos,</i> <i>Fábio Q.B. Da Silva, and Sérgio Soares</i>	
— <i>Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco;</i> <i>Universidade Federal de Pernambuco; Instituto Federal do Pará;</i> <i>Instituto Federal de Alagoas</i>	

Abnormal Working Hours: Effect of Rapid Releases and Implications to Work Content.....	243
<i>Maëlick Claes, Mika Mäntylä, Miikka Kuutila, and Bram Adams</i>	
— <i>University of Oulu; École Polytechnique de Montréal</i>	

## Clones and Edits

Mining Change Histories for Unknown Systematic Edits.....	248
<i>Tim Molderez, Reinout Stevens, and Coen De Roover</i>	
— <i>Vrije Universiteit Brussel</i>	
Source File Set Search for Clone-and-Own Reuse Analysis.....	257
<i>Takashi Ishio, Yusuke Sakaguchi, Kaoru Ito, and Katsuro Inoue</i>	
— <i>Osaka University</i>	
RefDiff: Detecting Refactorings in Version Histories .....	269
<i>Daniilo Silva and Marco Tulio Valente</i>	
— <i>Universidade Federal de Minas Gerais</i>	
Stack Overflow in Github: Any Snippets There?.....	280
<i>Di Yang, Pedro Martins, Vaibhav Saini, and Cristina Lopes</i>	
— <i>University of California, Irvine</i>	
Some from Here, Some from There: Cross-Project Code Reuse in GitHub .....	291
<i>Mohammad Gharehyazie, Baishakhi Ray, and Vladimir Filkov</i>	
— <i>University of California, Davis; University of Virginia</i>	
Exception Evolution in Long-Lived Java Systems .....	302
<i>Haidar Osman, Andrei Chis, Claudio Corrodi, Mohammad Ghafari,</i>	
<i>and Oscar Nierstrasz</i>	
— <i>University of Bern; Feenk GmbH</i>	

## Continuous Integration and Build

Do Not Trust Build Results at Face Value—An Empirical Study of 30 Million CPAN Builds.....	312
<i>Mahdis Zolfagharinia, Bram Adams, and Yann-Gaël Guéhéneuc</i>	
— <i>École Polytechnique de Montréal</i>	
An Empirical Analysis of the Docker Container Ecosystem on GitHub .....	323
<i>Jürgen Cito, Gerald Schermann, John Erik Wittern, Philipp Leitner, Sali Zumberi,</i>	
<i>and Harald C. Gall</i>	
— <i>University of Zurich; IBM T.J. Watson Research Center</i>	
How Open Source Projects Use Static Code Analysis Tools in Continuous Integration Pipelines .....	334
<i>Fiorella Zampetti, Simone Scalabrino, Rocco Oliveto, Gerardo Canfora,</i>	
<i>and Massimiliano Di Penta</i>	
— <i>University of Sannio; University of Molise</i>	
An Empirical Analysis of Build Failures in the Continuous Integration Workflows of Java-Based Open-Source Software .....	345
<i>Thomas Rausch, Waldemar Hummer, Philipp Leitner, and Stefan Schulte</i>	
— <i>Vienna University of Technology; University of Zurich</i>	

Oops, My Tests Broke the Build: An Explorative Analysis of Travis CI with GitHub ..... 356  
*Moritz Beller, Georgios Gousios, and Andy Zaidman*  
— *Delft University of Technology*

Extracting Build Changes with BUILDDIFF..... 368  
*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 ..... 379  
*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 ..... 391  
*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..... 402  
*Davide Spadini, Maurício Aniche, Magiel Bruntink, and Alberto Bacchelli*  
— *Software Improvement Group; Delft University of Technology*

Bug Characteristics in Blockchain Systems: A Large-Scale Empirical Study..... 413  
*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 ..... 425  
*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..... 436  
*Rana Alkadhi, Teodora Lata, Emitza Guzman, and Bernd Bruegge*  
— *Technische Universität München; University of Zurich*

## Mining Challenge

TravisTorrent: Synthesizing Travis CI and GitHub for Full-Stack Research  
on Continuous Integration ..... 447  
*Moritz Beller, Georgios Gousios, and Andy Zaidman*  
— *Delft University of Technology*

On the Differences between Unit and Integration Testing in the TravisTorrent Dataset ..... 451  
*Gerardo Orellana, Gulsher Laghari, Alessandro Murgia, and Serge Demeyer*  
— *University of Antwerp*

Cost-Effective Build Outcome Prediction Using Cascaded Classifiers.....	455
<i>Ansong Ni and Ming Li</i>	
— <i>Nanjing University</i>	
Sentiment Analysis of Travis CI Builds .....	459
<i>Rodrigo Souza and Bruno Silva</i>	
— <i>Salvador University; Federal University of Bahia</i>	
A Time Series Analysis of TravisTorrent Builds: To Everything There Is a Season.....	463
<i>Abigail Atchison, Christina Berardi, Natalie Best, Elizabeth Stevens,</i>	
<i>and Erik Linstead</i>	
— <i>Chapman University</i>	
Insights into Continuous Integration Build Failures.....	467
<i>Md Rakibul Islam and Minhaz F. Zibran</i>	
— <i>University of New Orleans</i>	
An Empirical Study of the Personnel Overhead of Continuous Integration.....	471
<i>Marco Manglaviti, Eduardo Coronado-Montoya, Keheliya Gallaba,</i>	
<i>and Shane McIntosh</i>	
— <i>McGill University</i>	
How Does Contributors' Involvement Influence the Build Status of an Open-Source Software Project?.....	475
<i>Marcel Rebouças, Renato O. Santos, Gustavo Pinto, and Fernando Castor</i>	
— <i>Federal University of Pernambuco; Federal Institute of Pará</i>	
On the Interplay between Non-Functional Requirements and Builds on Continuous Integration .....	479
<i>Klérisson V. R. Paixão, Crícia Z. Felício, Fernanda M. Delfim,</i>	
<i>and Marcelo De A. Maia</i>	
— <i>Universidade Federal de Uberlândia; Instituto Federal do Triângulo Mineiro</i>	
Analyzing the Impact of Social Attributes on Commit Integration Success .....	483
<i>Mauricio Soto, Zack Coker, and Claire Le Goues</i>	
— <i>Carnegie Mellon University</i>	
Built to Last or Built Too Fast? Evaluating Prediction Models for Build Times.....	487
<i>Ekaba Bisong, Eric Tran, and Olga Baysal</i>	
— <i>Carleton University</i>	
The Impact of the Adoption of Continuous Integration on Developer Attraction and Retention.....	491
<i>Yash Gupta, Yusaira Khan, Keheliya Gallaba, and Shane McIntosh</i>	
— <i>McGill University</i>	
An Empirical Study of Activity, Popularity, Size, Testing, and Stability in Continuous Integration .....	495
<i>Aakash Gautam, Saket Vishwasrao, and Francisco Servant</i>	
— <i>Virginia Tech</i>	
Impact of Continuous Integration on Code Reviews.....	499
<i>Mohammad Masudur Rahman, and Chanchal K. Roy</i>	
— <i>University of Saskatchewan</i>	



Prevalence of Botched Code Integrations .....	503
<i>Ward Muylaert and Coen De Roover</i>	
— <i>Vrije Universiteit Brussel</i>	
<b>Data Showcase</b>	
Software Evolution and Quality Data from Controlled, Multiple, Industrial Case Studies.....	507
<i>Aiko Yamashita, S. Amirhossein Abtahizadeh, Foutse Khomh,</i>	
<i>and Yann-Gaël Guéhéneuc</i>	
— <i>Centrum Wiskunde &amp; Informatica; École Polytechnique de Montréal</i>	
A Dataset of Scratch Programs: Scraped, Shaped and Scored .....	511
<i>Efthimia Aivaloglou, Felienne Hermans, Jesus Moreno-Leon,</i>	
<i>and Gregorio Robles</i>	
— <i>Delft University of Technology; Programamos.es;</i>	
<i>Universidad Rey Juan Carlos</i>	
Continuous Defect Prediction: The Idea and a Related Dataset .....	515
<i>Lech Madeyski and Marcin Kawalerowicz</i>	
— <i>Wroclaw University of Science and Technology; Opole University of Technology</i>	
An Extensive Dataset of UML Models in GitHub .....	519
<i>Gregorio Robles, Truong Ho-Quang, Regina Hebig, Michel R.V. Chaudron,</i>	
<i>and Miguel Angel Fernandez</i>	
— <i>Universidad Rey Juan Carlos; Chalmers &amp; Gothenburg University</i>	
A Dataset for Dynamic Discovery of Semantic Changes in Version Controlled Software Histories .....	523
<i>Chenguang Zhu, Yi Li, Julia Rubin, and Marsha Chechik</i>	
— <i>University of Toronto; University of British Columbia</i>	
Rediscovery Datasets: Connecting Duplicate Reports .....	527
<i>Mefta Sadat, Ayse Basar Bener, and Andriy Miranskyy</i>	
— <i>Ryerson University</i>	
A Data Set of OCL Expressions on GitHub .....	531
<i>Jeroen Noten, Josh G.M. Mengerink, and Alexander Serebrenik</i>	
— <i>Eindhoven University of Technology</i>	
<b>Author Index .....</b>	<b>535</b>