

2015 IEEE/ACM 12th Working Conference on Mining Software Repositories

(MSR 2015)

**Florence, Italy
16-17 May 2015**



**IEEE Catalog Number: CFP1578C-POD
ISBN: 978-1-4673-7924-3**

The 12th Working Conference on Mining Software Repositories (MSR 2015)

Table of Contents

Message from the Chairs	xii
Committees and Reviewers	xiv
Sponsors and Supporters	xviii

Keynote

Confessions of a Worldly Software Miner	1
<i>Radu Marinescu — Politehnica University of Timisoara, Romania</i>	

Practice Papers (Reports from the Trenches)

Code Ownership and Software Quality: A Replication Study	2
<i>Michaela Greiler, Kim Herzig, Jacek Czerwonka — Microsoft, USA; Microsoft Research, UK; Microsoft Corp., USA</i>	
Extracting Facts from Performance Tuning History of Scientific Applications for Predicting Effective Optimization Patterns	13
<i>Masatomo Hashimoto, Masaaki Terai, Toshiyuki Maeda, Kazuo Minami — RIKEN Advanced Institute for Computational Science, Japan</i>	
Mining Component Repositories for Installability Issues	24
<i>Pietro Abate, Roberto Di Cosmo, Louis Gesbert, Fabrice Le Fessant, Ralf Treinen, Stefano Zacchiroli — INRIA, France; Université Paris Diderot, France; OCamlPro, France</i>	

Everything Changes (Or Stays the Same)

The Uniqueness of Changes: Characteristics and Applications	34
<i>Baishakhi Ray, Meiyappan Nagappan, Christian Bird, Nachiappan Nagappan, Thomas Zimmermann — University of California at Davis, USA; Rochester Institute of Technology, USA; Microsoft Research, USA</i>	
Co-evolution of Infrastructure and Source Code - An Empirical Study	45
<i>Yujuan Jiang, Bram Adams — Polytechnique Montreal, Canada</i>	
Mining Energy-Aware Commits	56
<i>Irineu Moura, Gustavo Pinto, Felipe Ebert, Fernando Castor — Federal University of Pernambuco, Brazil</i>	

Why Power Laws? An Explanation from Fine-Grained Code Changes 68
Zhongpeng Lin, Jim Whitehead — University of California at Santa Cruz, USA

Sameness: An Experiment in Code Search..... 76
Lee Martie, André van der Hoek — University of California at Irvine, USA

Interaction Data and App Mining

Using Developer-Interaction Trails to Triage Change Requests 88
*Motahareh Bahrami Zanjani, Huzefa Kagdi, Christian Bird — Wichita State University, USA;
Microsoft Research, USA*

An Empirical Study of the Copy and Paste Behavior during Development 99
Tarek M. Ahmed, Weiyi Shang, Ahmed E. Hassan — Queen's University, Canada

Mining Android App Usages for Generating Actionable GUI-Based Execution Scenarios 111
*Mario Linares-Vásquez, Martin White, Carlos Bernal-Cárdenas, Kevin Moran,
Denys Poshyvanyk — College of William and Mary, USA*

The App Sampling Problem for App Store Mining 123
*William Martin, Mark Harman, Yue Jia, Federica Sarro, Yuanyuan Zhang
— University College London, UK*

Unveiling Exception Handling Bug Hazards in Android Based on GitHub and Google
Code Issues 134
*Roberta Coelho, Lucas Almeida, Georgios Gousios, Arie van Deursen — Federal University
of Rio Grande do Norte, Brazil; Radboud University Nijmegen, Netherlands; Delft University
of Technology, Netherlands*

Code Review (That Passed Peer Review)

Characteristics of Useful Code Reviews: An Empirical Study at Microsoft 146
*Amiangshu Bosu, Michaela Greiler, Christian Bird — University of Alabama, USA;
Microsoft Research, USA*

Will They Like This? Evaluating Code Contributions with Language Models..... 157
*Vincent J. Hellendoorn, Premkumar T. Devanbu, Alberto Bacchelli — Delft University
of Technology, Netherlands; University of California at Davis, USA*

Investigating Code Review Practices in Defective Files: An Empirical Study of the Qt System 168
*Patanamon Thongtanunam, Shane McIntosh, Ahmed E. Hassan, Hajimu Iida
— Nara Institute of Science and Technology, Japan; Queen's University, Canada*

Partitioning Composite Code Changes to Facilitate Code Review..... 180
*Yida Tao, Sunghun Kim — The Hong Kong University of Science and Technology,
Hong Kong*

Lessons Learned from Building and Deploying a Code Review Analytics Platform..... 191
*Christian Bird, Trevor Carnahan, Michaela Greiler — Microsoft Research, USA;
Microsoft, USA; Microsoft, Germany*

Ecosystems, APIs, and Architecture

Ecosystems in GitHub and a Method for Ecosystem Identification Using Reference Coupling	202
<i>Kelly Blincoe, Francis Harrison, Daniela Damian — University of Victoria, Canada</i>	
A Historical Analysis of Debian Package Incompatibilities	212
<i>Maelick Claes, Tom Mens, Roberto Di Cosmo, Jérôme Vouillon — University of Mons, Belgium; Université Paris Diderot, France; INRIA, France</i>	
Recommending Posts concerning API Issues in Developer Q&A Sites	224
<i>Wei Wang, Haroon Malik, Michael W. Godfrey — University of Waterloo, Canada</i>	
An Empirical Study of Architectural Change in Open-Source Software Systems	235
<i>Duc Minh Le, Pooyan Behnamghader, Joshua Garcia, Daniel Link, Arman Shahbazian, Nenad Medvidovic — University of Southern California, USA; George Mason University, USA</i>	
A Study on the Role of Software Architecture in the Evolution and Quality of Software	246
<i>Ehsan Kouroshfar, Mehdi Mirakhorli, Hamid Bagheri, Lu Xiao, Sam Malek, Yuanfang Cai — George Mason University, USA; Rochester Institute of Technology, USA; Drexel University, USA</i>	

Scary Stuff: Bugs, Risks, and Vulnerabilities

Are These Bugs Really "Normal"?	258
<i>Ripon K. Saha, Julia Lawall, Sarfraz Khurshid, Dewayne E. Perry — University of Texas at Austin, USA; Sorbonne University, France</i>	
Do Bugs Foreshadow Vulnerabilities? A Study of the Chromium Project	269
<i>Felivel Camilo, Andrew Meneely, Meiyappan Nagappan — Rochester Institute of Technology, USA</i>	
Characterization and Prediction of Issue-Related Risks in Software Projects	280
<i>Morakot Choetkiertikul, Hoa Khanh Dam, Truyen Tran, Aditya Ghose — University of Wollongong, Australia; Deakin University, Australia</i>	

Computer Musicians Bullied for Using Gists

An Empirical Study of End-User Programmers in the Computer Music Community	292
<i>Gregory Bulet, Abram Hindle — University of Alberta, Canada</i>	
Are Bullies More Productive? Empirical Study of Affectiveness vs. Issue Fixing Time	303
<i>Marco Ortu, Bram Adams, Giuseppe Destefanis, Parastou Tourani, Michele Marchesi, Roberto Tonelli — University of Cagliari, Italy; École Polytechnique de Montréal, Canada; CRIM, Canada</i>	
What Is the Gist? Understanding the Use of Public Gists on GitHub	314
<i>Weiliang Wang, Germán Poo-Caamaño, Evan Wilde, Daniel M. German — University of Victoria, Canada</i>	

Licenses, Deep Learning, and Process Mining

A Method to Detect License Inconsistencies in Large-Scale Open Source Projects	324
<i>Yuhao Wu, Yuki Manabe, Tetsuya Kanda, Daniel M. German, Katsuro Inoue</i> — <i>Osaka University, Japan; Kumamoto University, Japan; University of Victoria, Canada</i>	
Toward Deep Learning Software Repositories	334
<i>Martin White, Christopher Vendome, Mario Linares-Vasquez, Denys Poshyvanyk</i> — <i>College of William and Mary, USA</i>	
Identifying Software Process Management Challenges: Survey of Practitioners in a Large Global IT Company	346
<i>Monika Gupta, Ashish Sureka, Srinivas Padmanabhuni, Allahbaksh Mohammedali Asadullah</i> — <i>Indraprastha Institute of Information Technology, India; Infosys Technologies Ltd., India</i>	

Short Papers

Automatically Prioritizing Pull Requests.....	357
<i>Erik van der Veen, Georgios Gousios, Andy Zaidman</i> — <i>Delft University of Technology, Netherlands; Radboud University Nijmegen, Netherlands</i>	
Matching GitHub Developer Profiles to Job Advertisements.....	362
<i>Claudia Hauff, Georgios Gousios</i> — <i>Delft University of Technology, Netherlands; Radboud University Nijmegen, Netherlands</i>	
Wait for It: Determinants of Pull Request Evaluation Latency on GitHub.....	367
<i>Yue Yu, Huaimin Wang, Vladimir Filkov, Premkumar Devanbu, Bogdan Vasilescu</i> — <i>National University of Defense Technology, China; University of California, USA</i>	
Toward Reusing Code Changes.....	372
<i>Yoshiki Higo, Akio Ohtani, Shinpei Hayashi, Hideaki Hata, Kusumoto Shinji</i> — <i>Osaka University, Japan; Tokyo Institute of Technology, Japan; Nara Institute of Science and Technology, Japan</i>	
Modifications, Tweaks, and Bug Fixes in Architectural Tactics	377
<i>Mehdi Mirakhorli, Jane Cleland-Huang</i> — <i>Rochester Institute of Technology, USA; DePaul University, USA</i>	
Do Onboarding Programs Work?.....	381
<i>Adriaan Labuschagne, Reid Holmes</i> — <i>University of Waterloo, Canada</i>	
An Enhanced Graph-Based Infrastructure for Software Search Engines.....	386
<i>Marcus Schumacher, Colin Atkinson</i> — <i>University of Mannheim, Germany</i>	
Organizational Volatility and Post-release Defects: A Replication Case Study Using Data from Google Chrome.....	391
<i>Samuel M. Donadelli, Yue Cai Zhu, Peter C. Rigby</i> — <i>Concordia University, Canada</i>	
Detecting and Mitigating Secret-Key Leaks in Source Code Repositories.....	396
<i>Vibha Singhal Sinha, Diptikalyan Saha, Pankaj Dhoolia, Rohan Padhye, Senthil Mani</i> — <i>IBM Research, India</i>	

Summarizing Complex Development Artifacts by Mining Heterogeneous Data.....	401
<i>Luca Ponzanelli, Andrea Mocchi, Michele Lanza — University of Lugano, Switzerland</i>	

Mining Challenge

The Synergy between Voting and Acceptance of Answers on StackOverflow - Or the Lack Thereof.....	406
<i>Neelamadhav Gantayat, Pankaj Dhoolia, Rohan Padhye, Senthil Mani, Vibha Singhal Sinha — IBM Research, India</i>	
Quality Questions Need Quality Code: Classifying Code Fragments on Stack Overflow	410
<i>Maarten Duijn, Adam Kucera, Alberto Bacchelli — Delft University of Technology, Netherlands; Czech Technical University in Prague, Czech Republic</i>	
ETA: Estimated Time of Answer Predicting Response Time in Stack Overflow	414
<i>Jeffrey Goderie, Brynjolfur Mar Georgsson, Bastiaan van Graafeiland, Alberto Bacchelli — Delft University of Technology, Netherlands</i>	
Going Green: An Exploratory Analysis of Energy-Related Questions.....	418
<i>Haroon Malik, Peng Zhao, Michael Godfrey — University of Waterloo, Canada</i>	
Mining StackOverflow to Filter Out Off-Topic IRC Discussion.....	422
<i>Shaiful Alam Chowdhury, Abram Hindle — University of Alberta, Canada</i>	
An Insight into the Unresolved Questions at Stack Overflow	426
<i>Mohammad Masudur Rahman, Chanchal K. Roy — University of Saskatchewan, Canada</i>	
Mining Successful Answers in Stack Overflow	430
<i>Fabio Calefato, Filippo Lanubile, Maria Concetta Marasciulo, Nicole Novielli — University of Bari, Italy</i>	
Quick Trigger on Stack Overflow: A Study of Gamification-Influenced Member Tendencies	434
<i>Yong Jin, Xin Yang, Raula Gaikovina Kula, Eunjong Choi, Katsuro Inoue, Hajimu Iida — Nara Institute of Science and Technology, Japan; Osaka University, Japan</i>	
Intuition vs. Truth: Evaluation of Common Myths about StackOverflow Posts.....	438
<i>Verena Honsel, Steffen Herbold, Jens Grabowski — University of Göttingen, Germany</i>	
Automatic Assessments of Code Explanations: Predicting Answering Times on Stack Overflow	442
<i>Selman Ercan, Quinten Stokkink, Alberto Bacchelli — Delft University of Technology, Netherlands</i>	
Which Non-functional Requirements Do Developers Focus On? An Empirical Study on Stack Overflow Using Topic Analysis	446
<i>Jie Zou, Ling Xu, Weikang Guo, Meng Yan, Dan Yang, Xiaohong Zhang — Chongqing University, China</i>	
Stack Overflow Badges and User Behavior: An Econometric Approach.....	450
<i>Andrew Marder — Harvard Business School, USA</i>	
Employing Source Code Information to Improve Question-Answering in Stack Overflow	454
<i>Themistoklis Diamantopoulos, Andreas Symeonidis — Aristotle University of Thessaloniki, Greece</i>	

One-Day Flies on StackOverflow - Why the Vast Majority of StackOverflow Users Only Posts Once.....	458
<i>Rogier Slag, Mike de Waard, Alberto Bacchelli — Delft University of Technology, Netherlands</i>	

Data Showcase

A Repository with 44 Years of Unix Evolution.....	462
<i>Diomidis Spinellis — Athens University of Economics and Business, Greece</i>	
The Debsources Dataset: Two Decades of Debian Source Code Metadata	466
<i>Stefano Zacchiroli — University Paris Diderot, France</i>	
A Dataset of the Activity of the Git Super-repository of Linux in 2012.....	470
<i>Daniel M. German, Bram Adams, Ahmed E. Hassan — University of Victoria, Canada; École Polytechnique de Montréal, Canada; Queen's University, Canada</i>	
StORMeD: Stack Overflow Ready Made Data	474
<i>Luca Ponzanelli, Andrea Mocchi, Michele Lanza — University of Lugano, Switzerland</i>	
The MetricsGrimoire Database Collection	478
<i>Jesus M. Gonzalez-Barahona, Gregorio Robles, Daniel Izquierdo-Cortazar — Universidad Rey Juan Carlos, Spain; Bitergia, Spain</i>	
Landfill: An Open Dataset of Code Smells with Public Evaluation	482
<i>Fabio Palomba, Dario Di Nucci, Michele Tufano, Gabriele Bavota, Rocco Oliveto, Denys Poshyvanyk, Andrea De Lucia — University of Salerno, Italy; College of William and Mary, USA; Free University of Bolzano-Bozen, Italy; University of Molise, Italy</i>	
Fuse: A Reproducible, Extendable, Internet-Scale Corpus of Spreadsheets.....	486
<i>Titus Barik, Kevin Lubick, Justin Smith, John Slankas, Emerson Murphy-Hill — North Carolina State University, USA</i>	
Dataset of Developer-Labeled Commit Messages	490
<i>Andreas Mauczka, Florian Brosch, Christian Schanes, Thomas Grechenig — Vienna University of Technology, Austria</i>	
A Novel Industry Grade Dataset for Fault Prediction Based on Model-Driven Developed Automotive Embedded Software	494
<i>Harald Altinger, Sebastian Siegl, Yanja Dajsuren, Franz Wotawa — Audi Electronics Venture GmbH, Germany; Eindhoven University of Technology, Netherlands; Technische Universitaet Graz, Austria</i>	
The Firefox Temporal Defect Dataset.....	498
<i>Mayy Habayeb, Andriy Miranskyy, Syed Shariyar Murtaza, Leotis Buchanan, Ayse Bener Ryerson University, Canada</i>	
An Architectural Evolution Dataset	502
<i>Michel Wermelinger, Yijun Yu — The Open University, UK</i>	
A Dataset for API Usage	506
<i>Anand Ashok Sawant, Alberto Bacchelli — Delft University of Technology, Netherlands</i>	

Generating the Blueprints of the Java Ecosystem	510
<i>Vassilios Karakoidas, Dimitrios Mitropoulos, Panos Louridas, Georgios Gousios, Diomidis Spinellis — Athens University of Economics and Business, Greece; Columbia University, USA; Radboud University Nijmegen, Netherlands</i>	
A Data Set for Social Diversity Studies of GitHub Teams	514
<i>Bogdan Vasilescu, Alexander Serebrenik, Vladimir Filkov — University of California at Davis, USA; Eindhoven University of Technology, Netherlands</i>	
A Dataset of High Impact Bugs: Manually-Classified Issue Reports	518
<i>Masao Ohira, Yutaro Kashiwa, Yosuke Yamatani, Hayato Yoshiyuki, Yoshiya Maeda, Nachai Limsettho, Keisuke Fujino, Hideaki Hata, Akinori Ihara, Kenichi Matsumoto — Wakayama University, Japan; Nara Institute of Science and Technology, Japan</i>	
A Dataset of Open-Source Android Applications	522
<i>Daniel E. Krutz, Mehdi Mirakhorli, Samuel A. Malachowsky, Andres Ruiz, Jacob Peterson, Andrew Filipski, Jared Smith — Rochester Institute of Technology, USA</i>	
Author Index	526