# 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  
*Radu Marinescu — Politehnica University of Timisoara, Romania*

## Practice Papers (Reports from the Trenches)

- Code Ownership and Software Quality: A Replication Study ................................................................. 2  
  *Michaela Greiler, Kim Herzig, Jacek Czerwonka — Microsoft, USA; Microsoft Research, UK;  
  Microsoft Corp., USA*

- Extracting Facts from Performance Tuning History of Scientific Applications for Predicting Effective Optimization Patterns ................................................................................................................... 13  
  *Masatomo Hashimoto, Masaaki Terai, Toshiyuki Maeda, Kazuo Minami — RIKEN Advanced Institute for Computational Science, Japan*

- Mining Component Repositories for Installability Issues ............................................................................ 24  
  *Pietro Abate, Roberto Di Cosmo, Louis Gesbert, Fabrice Le Fessant, Ralf Treinen,  
  Stefano Zacchioli — INRIA, France; Université Paris Diderot, France; OCamlPro, France*

## Everything Changes (Or Stays the Same)

- The Uniqueness of Changes: Characteristics and Applications ................................................................. 34  
  *Baishakhi Ray, Meiyappan Nagappan, Christian Bird, Nachiappan Nagappan,  
  Thomas Zimmermann — University of California at Davis, USA; Rochester Institute of Technology, USA; Microsoft Research, USA*

- Co-evolution of Infrastructure and Source Code - An Empirical Study ....................................................... 45  
  *Yujuan Jiang, Bram Adams — Polytechnique Montreal, Canada*

- Mining Energy-Aware Commits ................................................................................................................... 56  
  *Irineu Moura, Gustavo Pinto, Felipe Ebert, Fernando Castor — Federal University of Pernambuco, Brazil*
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, Wei Yi 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

   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
Kelly Blincoe, Francis Harrison, Daniela Damian — University of Victoria, Canada

A Historical Analysis of Debian Package Incompatibilities ................................................................. 212
Maelick Claes, Tom Mens, Roberto Di Cosmo, Jérôme Vouillon — University of Mons, Belgium; Université Paris Diderot, France; INRIA, France

Recommending Posts concerning API Issues in Developer Q&A Sites .................................................. 224
Wei Wang, Haroon Malik, Michael W. Godfrey — University of Waterloo, Canada

An Empirical Study of Architectural Change in Open-Source Software Systems ............................... 235
Duc Minh Le, Pooyan Behnamghader, Joshua Garcia, Daniel Link, Arman Shahbazian, Nenad Medvidovic — University of Southern California, USA; George Mason University, USA

A Study on the Role of Software Architecture in the Evolution and Quality of Software .................. 246
Ehsan Kouroshfar, Mehdi Mirakhorli, Hamid Bagheri, Lu Xiao, Sam Malek, Yuanfang Cai — George Mason University, USA; Rochester Institute of Technology, USA; Drexel University, USA

Scary Stuff: Bugs, Risks, and Vulnerabilities

Are These Bugs Really "Normal"? ............................................................................................................ 258
Ripon K. Saha, Julia Lawall, Sarfraz Khurshid, Dewayne E. Perry — University of Texas at Austin, USA; Sorbonne University, France

Do Bugs Foreshadow Vulnerabilities? A Study of the Chromium Project ........................................ 269
Felivel Camilo, Andrew Meneely, Meiyappan Nagappan — Rochester Institute of Technology, USA

Characterization and Prediction of Issue-Related Risks in Software Projects ....................................... 280
Morakot Choetkiertikul, Hoa Khanh Dam, Truyen Tran, Aditya Ghose — University of Wollongong, Australia; Deakin University, Australia

Computer Musicians Bullied for Using Gists

An Empirical Study of End-User Programmers in the Computer Music Community .......................... 292
Gregory Burlet, Abram Hindle — University of Alberta, Canada

Marco Ortu, Bram Adams, Giuseppe Destefanis, Parastou Tourani, Michele Marchesi, Roberto Tonelli — University of Cagliari, Italy; École Polytechnique de Montréal, Canada; CRIM, Canada

What Is the Gist? Understanding the Use of Public Gists on GitHub ................................................ 314
Weiliang Wang, Germán Poo-Caamaño, Evan Wilde, Daniel M. German — University of Victoria, Canada
Licenses, Deep Learning, and Process Mining

A Method to Detect License Inconsistencies in Large-Scale Open Source Projects ........................................ 324
Yuhao Wu, Yuki Manabe, Tetsuya Kanda, Daniel M. German, Katsuro Inoue
— Osaka University, Japan; Kumamoto University, Japan; University of Victoria, Canada

Toward Deep Learning Software Repositories ..................................................................................................... 334
Martin White, Christopher Vendome, Mario Linares-Vasquez, Denys Poshyvanyk
— College of William and Mary, USA

Identifying Software Process Management Challenges: Survey of Practitioners in a Large
Global IT Company ............................................................................................................................................... 346
Monika Gupta, Ashish Sureka, Srinivas Padmanabhuni, Allahbakhsh Mohammed Ali Asadullah
— Indraprastha Institute of Information Technology, India; Infosys Technologies Ltd., India

Short Papers

Automatically Prioritizing Pull Requests ........................................................................................................... 357
Erik van der Veen, Georgios Gousios, Andy Zaidman — Delft University of Technology,
Netherlands; Radboud University Nijmegen, Netherlands

Matching GitHub Developer Profiles to Job Advertisements ........................................................................ 362
Claudia Hauff, Georgios Gousios — Delft University of Technology, Netherlands;
Radboud University Nijmegen, Netherlands

Wait for It: Determinants of Pull Request Evaluation Latency on GitHub ..................................................... 367
Yue Yu, Huaizmin Wang, Vladimir Filkov, Premkumar Devanbu, Bogdan Vasilescu
— National University of Defense Technology, China; University of California, USA

Toward Reusing Code Changes ......................................................................................................................... 372
Yoshiki Higo, Akio Ohtani, Shinpei Hayashi, Hideaki Hata, Kusumoto Shinji
— Osaka University, Japan; Tokyo Institute of Technology, Japan; Nara Institute
of Science and Technology, Japan

Modifications, Tweaks, and Bug Fixes in Architectural Tactics .................................................................. 377
Mehdi Mirakhorli, Jane Cleland-Huang — Rochester Institute of Technology, USA;
DePaul University, USA

Do Onboarding Programs Work? ................................................................................................................... 381
Adriaan Labuschagne, Reid Holmes — University of Waterloo, Canada

An Enhanced Graph-Based Infrastructure for Software Search Engines .................................................... 386
Marcus Schumacher, Colin Atkinson — University of Mannheim, Germany

Organizational Volatility and Post-release Defects: A Replication Case Study Using Data
from Google Chrome ........................................................................................................................................... 391
Samuel M. Donadelli, Yue Cai Zhu, Peter C. Rigby — Concordia University, Canada

Detecting and Mitigating Secret-Key Leaks in Source Code Repositories .................................................. 396
Vibha Singhal Sinha, Dipak Gyan Saha, Pankaj Dhoolia, Rohan Padhye, Senthil Mani
— IBM Research, India
Summarizing Complex Development Artifacts by Mining Heterogeneous Data

Luca Ponzanelli, Andrea Mocci, Michele Lanza — University of Lugano, Switzerland

**Mining Challenge**

The Synergy between Voting and Acceptance of Answers on StackOverflow

- Or the Lack Thereof

Neelamadhav Gantayat, Pankaj Dhoolia, Rohan Padhye, Senthil Mani, Vibha Singhal Sinha

— IBM Research, India

Quality Questions Need Quality Code: Classifying Code Fragments on Stack Overflow

Maarten Duijn, Adam Kucera, Alberto Bacchelli — Delft University of Technology, Netherlands; Czech Technical University in Prague, Czech Republic

ETA: Estimated Time of Answer Predicting Response Time in Stack Overflow

Jeffrey Goderie, Brynjolfur Mar Georgsson, Bastiaan van Graafeiland, Alberto Bacchelli

— Delft University of Technology, Netherlands

Going Green: An Exploratory Analysis of Energy-Related Questions

Haroon Malik, Peng Zhao, Michael Godfrey — University of Waterloo, Canada

Mining StackOverflow to Filter Out Off-Topic IRC Discussion

Shaiful Alam Chowdhury, Abram Hindle — University of Alberta, Canada

An Insight into the Unresolved Questions at Stack Overflow

Mohammad Masudur Rahman, Chanchal K. Roy — University of Saskatchewan, Canada

Mining Successful Answers in Stack Overflow

Fabio Calefato, Filippo Lanubile, Maria Concetta Marasciulo, Nicole Novielli

— University of Bari, Italy

Quick Trigger on Stack Overflow: A Study of Gamification-Influenced Member Tendencies

Yong Jin, Xin Yang, Raula Gaikovina Kula, Eunjong Choi, Katsuro Inoue, Hajimu Iida

— Nara Institute of Science and Technology, Japan; Osaka University, Japan

Intuition vs. Truth: Evaluation of Common Myths about StackOverflow Posts

Verena Honsel, Steffen Herbold, Jens Grabowski — University of Göttingen, Germany

Automatic Assessments of Code Explanations: Predicting Answering Times on Stack Overflow

Selman Ercan, Quinten Stokkink, Alberto Bacchelli — Delft University of Technology, Netherlands

Which Non-functional Requirements Do Developers Focus On? An Empirical Study on Stack Overflow Using Topic Analysis

Jie Zou, Ling Xu, Weikang Guo, Meng Yan, Dan Yang, Xiaohong Zhang

— Chongqing University, China

Stack Overflow Badges and User Behavior: An Econometric Approach

Andrew Marder — Harvard Business School, USA

Employing Source Code Information to Improve Question-Answering in Stack Overflow

Themistoklis Diamantopoulos, Andreas Symeonidis — Aristotle University of Thessaloniki, Greece
One-Day Flies on StackOverflow - Why the Vast Majority of StackOverflow Users Only Posts Once

Rogier Slag, Mike de Waard, Alberto Bacchelli — Delft University of Technology, Netherlands

Data Showcase

A Repository with 44 Years of Unix Evolution

Diomidis Spinellis — Athens University of Economics and Business, Greece

The Debsources Dataset: Two Decades of Debian Source Code Metadata

Stefano Zacchioli — University Paris Diderot, France

A Dataset of the Activity of the Git Super-repository of Linux in 2012

Daniel M. German, Bram Adams, Ahmed E. Hassan — University of Victoria, Canada; École Polytechnique de Montréal, Canada; Queen’s University, Canada

StORMeD: Stack Overflow Ready Made Data

Luca Ponzanelli, Andrea Mocci, Michele Lanza — University of Lugano, Switzerland

The MetricsGrimoire Database Collection

Jesus M. Gonzalez-Barahona, Gregorio Robles, Daniel Izquierdo-Cortazar — Universidad Rey Juan Carlos, Spain; Bitergia, Spain

Landfill: An Open Dataset of Code Smells with Public Evaluation

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

Fuse: A Reproducible, Extendable, Internet-Scale Corpus of Spreadsheets

Titus Barik, Kevin Lubick, Justin Smith, John Slankas, Emerson Murphy-Hill — North Carolina State University, USA

Dataset of Developer-Labeled Commit Messages

Andreas Mauczka, Florian Brosch, Christian Schanes, Thomas Grechenig — Vienna University of Technology, Austria

A Novel Industry Grade Dataset for Fault Prediction Based on Model-Driven Developed Automotive Embedded Software

Harald Altinger, Sebastian Siegl, Yanja Dajsuren, Franz Wotawa — Audi Electronics Venture GmbH, Germany; Eindhoven University of Technology, Netherlands; Technische Universitaet Graz, Austria

The Firefox Temporal Defect Dataset

Mayy Habayeb, Andriy Miranskyy, Syed Shariyar Murtaza, Leotis Buchanan, Ayse Bener — Ryerson University, Canada

An Architectural Evolution Dataset

Michel Wermelinger, Yijun Yu — The Open University, UK

A Dataset for API Usage

Anand Ashok Sawant, Alberto Bacchelli — Delft University of Technology, Netherlands
Generating the Blueprints of the Java Ecosystem
Vassilios Karakoidas, Dimitrios Mitropoulos, Panos Louridas, Georgios Gousios, Diomidis Spinellis — Athens University of Economics and Business, Greece; Columbia University, USA; Radboud University Nijmegen, Netherlands

A Data Set for Social Diversity Studies of GitHub Teams
Bogdan Vasilescu, Alexander Serebrenik, Vladimir Filkov — University of California at Davis, USA; Eindhoven University of Technology, Netherlands

A Dataset of High Impact Bugs: Manually-Classified Issue Reports
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

A Dataset of Open-Source Android Applications
Daniel E. Krutz, Mehdi Mirakhorli, Samuel A. Malachowsky, Andres Ruiz, Jacob Peterson, Andrew Filipski, Jared Smith — Rochester Institute of Technology, USA

Author Index