



Program of the Twelfth Working Conference on Mining Software Repositories

— Saturday, May 16 —

📌 Distinguished Paper | 📄 Journal Invitee | 🐦 Twitter: #msr15 | 📘 Facebook: facebook.com/msrconf

MSR 2015 Opening

Sat, May 16, 08:15 – 08:30

Room: Sala Verde

Massimiliano Di Penta, Martin Pinzger, Romain Robbes, Annie Ying & Yasutaka Kamei

Keynote

Sat, May 16, 08:30 – 09:30

Room: Sala Verde

Session Chair: Massimiliano Di Penta

Confessions of a Worldly Software Miner?

Radu Marinescu (Politehnica University of Timisoara, Romania)



Once upon a time I was an idealistic researcher. I started mining software in order to spot and correct design flaws. Like any idealistic researcher, I was creating "approaches", building "prototypes", performing "empirical studies", writing papers and giving talks; and I was happy whenever my work was appreciated by my fellows. I was living in a wonderful world (except for funding). Nevertheless, I heard rumors that out there lies another world: the software industry. I was confident that this other world is desperately waiting for my wonderful approaches. I was wrong. I found a world that is indeed seeking for new ways to improve quality and productivity; but it's a very hostile environment: "prototypes" crash on most real systems, academic "approaches" are nice but way too imprecise, and most "empirical studies" provide answers to questions that nobody seems to ask. I soon realized that the difference is deeper than "just engineering": the quality problems faced by real world projects are radically different from those that I read and wrote about in my idealistic "in vitro" environment: they are uglier, more complex, and amplified by uncountable technical and social constraints. I faced a tough decision: take the "blue pill" and return to my wonderful world, or take the "red pill" and continue to face the ugly reality. I chose the latter. In this talk I will share the most important lessons learned so far from my ongoing quest to put at work the best available code analysis techniques and tools for solving real quality issues. Some lessons may be rough and irreverent, but aren't most real life lessons like that?

Radu Marinescu is a professor of software engineering at the Politehnica University of Timisoara, Romania, where he heads the LOOSE Research Group. His research is focused on the areas of quality assurance, software metrics and refactoring. He strongly believes that research must ultimately flow into software products that will change the state of the practice in software companies. That's why he co-founded Intooituis, a spin-off company that provides disruptive tools reflecting a novel approach on assessing design quality and detecting symptoms of technical debt. These tools are currently used by thousands of developers and architects in large-scale companies, including a very successful deployment in a Global Fortune 500 telecom company. In 2014 he received the ICSME Most Influential Paper Award, after having received in 2009 the IBM John Backus Award for "having done the most to improve programmer productivity" from a jury that included two Turing Award winners. In 2010 he served General Chair of ICSM.

Practice Papers (Reports from the Trenches)

Sat, May 16, 09:30 – 10:30

Room: Sala Verde

Session Chair: Christian Bird

Code Ownership and Software Quality: A Replication Study

Michaela Greiler, Kim Herzig and Jacek Czerwonka (Microsoft, United States, Microsoft Research, United Kingdom and Microsoft Corp., United States)

Extracting Facts from Performance Tuning History of Scientific Applications for Predicting Effective Optimization Patterns

Masatomo Hashimoto, Masaaki Terai, Toshiyuki Maeda and Kazuo Minami (RIKEN Advanced Institute for Computational Science, Japan)

Mining Component Repositories for Installability Issues

Pietro Abate, Roberto Di Cosmo, Louis Gesbert, Fabrice Le Fessant, Ralf Treinen and Stefano Zacchiroli (INRIA, France and Université Paris Diderot, France)

Coffee Break

Sat, May 16, 10:30 – 11:00

Everything Changes (or stays the Same)

Sat, May 16, 11:00 – 12:30

Room: Sala Verde

Session Chair: Denys Poshyvanyk

The Uniqueness of Changes: Characteristics and Applications

Baishakhi Ray, Meiyappan Nagappan, Christian Bird, Nachiappan Nagappan and Thomas Zimmermann (University of California, Davis, United States, Rochester Institute of Technology, United States and Microsoft Research, United States)

Co-evolution of infrastructure and Source Code - An Empirical Study

Yujuan Jiang and Bram Adams (Polytechnique Montreal, Canada)

Mining Energy-Aware Commits

Irineu Moura, Gustavo Pinto, Felipe Ebert and Fernando Castor (Federal University of Pernambuco, Brazil)

Why Power Laws? An Explanation from Fine-Grained Code Changes

Zhongpeng Lin and Jim Whitehead (University of California, Santa Cruz)

Sameness: An Experiment in Code Search
Lee Martie and André van der Hoek (UCI, United States)

Lunch
Sat, May 16, 12:30 – 14:00

Interaction Data and App Mining
Sat, May 16, 14:00 – 15:30

Room: Sala Verde

Session Chair: Mei Nagappan

Using Developer-Interaction Trails to Triage Change Requests

Motahareh Bahrami Zanjani, Huzefa Kagdi and Christian Bird (Wichita State University, United States and Microsoft Research, United States)

Studying Developers Copy and Paste Behavior

Tarek Ahmed, Weiyi Shang and Ahmed Hassan (Queen's University, Canada)

Mining Android App Usages for Generating Actionable GUI-based Execution Scenarios

Mario Linares-Vásquez, Martin White, Carlos Eduardo Bernal Cardenas, Kevin Moran and Denys Poshyvanyk (The College of William and Mary, United States)

The App Sampling Problem for App Store Mining

William Martin, Mark Harman, Yue Jia, Federica Sarro and Yuanyuan Zhang (University College London, United Kingdom)

Unveiling Exception Handling Bug Hazards in Android based on GitHub and Google Code Issues

Roberta Coelho, Lucas Almeida, Georgios Gousios and Arie van Deursen (Federal University of Rio Grande do Norte, Brazil, Radboud University Nijmegen, Netherlands and Delft University of Technology, Netherlands)

Coffee Break
Sat, May 16, 15:30 – 16:00

MSR Challenge; 14 Challenge Presentations
Sat, May 16, 16:00 – 17:00

Room: Sala Verde

Session Chair: Laura Moreno

The Synergy Between Voting and Acceptance of Answers on StackOverflow, or the Lack thereof

Neelamadhav Gantayat, Pankaj Dhoolia, Rohan Padhye, Senthil Mani and Vibha Singhal Sinha (IBM Research, India)

Quality questions need quality code: Classifying code fragments on StackOverflow

Maarten Duijn, Adam Kucera and Alberto Bacchelli (Delft University of Technology, Netherlands and 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 and Alberto Bacchelli (Delft University of Technology, Netherlands)

Going Green: An Exploratory Analysis of Energy-Related Questions

Haroon Malik, Peng Zhao and Michael Godfrey (University of Waterloo, Canada)

Mining StackOverflow to Filter out Off-topic IRC Discussion

Shaiful Chowdhury and Abram Hindle (University of Alberta, Canada)

An Insight into the Unresolved Questions at Stack Overflow

Mohammad Masudur Rahman and Chanchal K. Roy (University of Saskatchewan, Canada)

Mining Successful Answers in Stack Overflow

Fabio Calefato, Filippo Lanubile, Maria Concetta Marasciulo and 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, Hajimu Iida and Katsuro Inoue (Nara Institute of Science and Technology and Osaka University, Japan)

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

Verena Honsel, Steffen Herbold and Jens Grabowski (Universität Göttingen, Germany)

Automatic Assessments of Code Explanations: Predicting answering times on Stack Overflow

Selman Ercan, Quinten Stokkink and 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 and Xiaohong Zhang (Chongqing University, China)

Stack Overflow badges and user behavior: An econometric approach

Andrew Marder (Harvard Business School, United States)

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

Themistoklis Diamantopoulos and 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 and Alberto Bacchelli (Delft University of Technology, Netherlands)

MSR Inaugural Parallel Sessions
Sat, May 16, 17:00 – 18:00

Data Papers

Room: Sala Verde

Session Chair: Yasutaka Kamei

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 Zacchiroli (Université Paris Diderot, France)

A Dataset of the Activity of the git superrepository of Linux

Daniel German, Bram Adams and Ahmed E. Hassan (University of Victoria, Canada, École Polytechnique de Montréal, Canada and Queen's University, Canada)

StORMeD: Stack Overflow Ready Made Data

Luca Ponzanelli, Andrea Mocci and Michele Lanza (University of Lugano, Switzerland)

The MetricsGrimoire Database Collection

Jesus M. Gonzalez-Barahona, Gregorio Robles and Daniel Izquierdo-Cortazar (Universidad Rey Juan Carlos, Spain, Universidad Rey Juan Carlos, Spain and 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 and Andrea De Lucia (University of Salerno, Italy, The College of William and Mary, United States, Free University of Bolzano-Bozen, Italy and University of Molise, Italy)

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

Titus Barik, Kevin Lubick, Justin Smith, John Slankas and Emerson Murphy-Hill (North Carolina State University, United States)

Dataset of developer-labeled commit messages for task classification validation

Andreas Mauczka, Florian Brosch, Christian Schanes and 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 and Franz Wotawa (Audi Electronics Venture GmbH, Germany, Audi Electronics Venture GmbH, Germany, Eindhoven University of Technology, Netherlands and Technische Universitaet Graz, Austria)

The Firefox Defect Temporal Dataset

Mayy Habayeb, Andriy Miranskyy, Syed Shariyar Murtaza, Leotis Buchanan and Ayse Bener (Ryerson University, Canada)

An Architectural Evolution Dataset

Michel Wermelinger and Yijun Yu (The Open University, United Kingdom)

A Dataset For API Usage

Anand Sawant and Alberto Bacchelli (Delft University of Technology, Netherlands)

Generating the Blueprints of the Java Ecosystem

Vassilios Karakoidas, Dimitris Mitropoulos, Georgios Gousios, Diomidis Spinellis and Panagiotis Louridas (Athens University of Economics and Business, Greece, Columbia University, United States and Radboud University Nijmegen, Netherlands)

A Data Set for Social Diversity Studies of GitHub Teams

Bogdan Vasilescu, Alexander Serebrenik and Vladimir Filkov (University of California, Davis, United States and 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 and Kenichi Matsumoto (Wakayama University, Japan and Nara Institute of Science and Technology, Japan)

A Dataset of Open Source Android Applications

Daniel Krutz, Mehdi Mirakhorli, Sam Malachowsky, Andres Ruiz, Jacob Peterson and Andrew Filipiski (Rochester Institute of Technology, United States)

Short Papers

Room: 203

Session Chair: Hongyu Zhang

Automatically Prioritizing Pull Requests

Erik van der Veen, Georgios Gousios and Andy Zaidman (Delft University of Technology, Netherlands and Radboud University Nijmegen, Netherlands)

Matching GitHub developer profiles to job advertisements

Claudia Hauff and Georgios Gousios (Delft University of Technology, Netherlands and Radboud University Nijmegen, Netherlands)

Wait For It: Determinants of Pull Request Evaluation Latency on GitHub

Yue Yu, Huaimin Wang, Vladimir Filkov, Premkumar Devanbu and Bogdan Vasilescu (National University of Defense Technology, China and University of California, United States)

Toward Reusing Code Changes

Yoshiki Higo, Akio Ohtani, Shinpei Hayashi, Hideaki Hata and Shinji Kusumoto (Osaka University, Japan and Tokyo Institute of Technology, Japan)

Modifications, Tweaks, and Bug Fixes in Architectural Tactics

Mehdi Mirakhorli and Jane Cleland-Huang (Rochester Institute of Technology, United States and DePaul, United States)

Do Onboarding Programs Work?

Adriaan Labuschagne and Reid Holmes (University of Waterloo, Canada)

An enhanced Graph-based infrastructure for Software Search Engines

Colin Atkinson and Marcus Schumacher (University of Mannheim, Germany)

Organizational volatility and post-release defects: A replication case study using data from Google Chrome

Samuel Mugnaini Donadelli, Yue Cai Zhu and Peter Rigby (Concordia University, Canada)

Detecting and Mitigating Secret-Key Leaks in Source Code Repositories

Vibha Singhal Sinha, Diptikalyan Saha, Pankaj Dhoolia, Rohan Padhye and Senthil Mani (IBM Research, India)

Summarizing Complex Development Artifacts by Mining Heterogenous Data

Luca Ponzanelli, Andrea Mocci and Michele Lanza (University of Lugano, Switzerland)

Conference Dinner and Awards Sat, May 16, TODO

Casa Trattoria

— Sunday, May 17 —

Announcements and MIP talk

Sun, May 17, 08:30 – 09:00

Room: Sala Verde

MSR Most Influential Paper Award
Distinguished paper awards
MSR Challenge winners, and Best data showcase award

Code Review (that passed Peer Review)

Sun, May 17, 09:00 – 10:30

Room: Sala Verde

Session Chair: Peter Rigby

Characteristics of Useful Code Reviews: An Empirical Study at Microsoft

Amiangshu Bosu, Michaela Greiler and Christian Bird (University of Alabama, United States, Microsoft Research, United States)

Will they like this? Evaluating Code Contributions With Language Models

Vincent Hellendoorn, Premkumar Devanbu and Alberto Bacchelli (Delft University of Technology, Netherlands and University of California, Davis)

Investigating Code Review Practices in Defective Files: An Empirical Study of the Qt System

Patanamon Thongtanunam, Shane McIntosh, Ahmed E. Hassan and Hajimu Iida (Nara Institute of Science and Technology, Japan and Queen's University, Canada)

Partitioning Composite Code Changes to Facilitate Code Review

Yida Tao and Sunghun Kim (The Hong Kong University of Science and Technology, Hong Kong)

Lessons Learned from Building and Deploying a Code Review Analytics Platform

Christian Bird, Trevor Carnahan and Micheala Greiler (Microsoft Research, United States, Microsoft, United States and Microsoft, Germany)

Coffee Break

Sun, May 17, 10:30 – 11:00

Ecosystems, APIs, and Architecture

Sun, May 17, 11:00 – 12:30

Room: Sala Verde

Session Chair: Andrew Begel

Ecosystems in GitHub and a Method for Ecosystem Identification using Reference Coupling

Kelly Blincoe, Francis Harrison and Daniela Damian (University of Victoria, New Zealand, SEGAL Labs, Canada and University of Victoria, Canada)

A historical analysis of Debian package incompatibilities

Tom Mens, Mailick Claes, Roberto Di Cosmo and Jerome Vouillon (University of Mons, Belgium, Université Paris Diderot, France and INRIA, France)

Recommending Posts Concerning API Issues in Developer Q&A Sites

Wei Wang, Haroon Malik and Mike Godfrey (University of Waterloo, Canada)

An Empirical Study of Architectural Change in Open-Source Software Systems

Duc Le, Pooyan Behnamghader, Joshua Garcia, Daniel Link, Arman Shahbazian and Nenad Medvidovic (University of Southern California, United States and George Mason University, United States)

A Study on the Role of Software Architecture in the Evolution and Quality of Software

Ehsan Kouroshfar, Mehdi Mirakhorli, Hamid Bagheri, Lu Xiao, Sam Malek and Yuanfang Cai (George Mason University, United States, Rochester Institute of Technology, United States and Drexel University, United States)

Lunch

Sun, May 17, 12:30 – 13:30

Data and Short Papers Poster Session

Sun, May 17, 13:30 – 14:30

Room: Sala Verde

Session Chair: TODO

Scary stuff: Bugs, Risks, and Vulnerabilities

Sun, May 17, 14:30 – 15:30

Room: Sala Verde

Session Chair: Bram Adams

Are These Bugs Really 'Normal'?

Ripon Saha, Julia Lawall, Sarfraz Khurshid and Dewayne E. Perry (The University of Texas at Austin, United States and Sorbonne University, France)

Do Bugs Foreshadow Vulnerabilities? A Study of the Chromium Project

Felivel Camilo, Andrew Meneely and Meiyappan Nagappan (Rochester Institute of Technology, United States)

Characterization and prediction of issue-related risks in software projects

Morakot Choetkiertikul, Hoa Khanh Dam, Truyen Tran and Aditya Ghose (University of Wollongong, Australia and Deakin University, Australia)

Coffee Break

Sun, May 17, 15:30 – 16:00

Computer Musicians Bullied for Using Gists

Sun, May 17, 16:00 – 17:00

Room: Sala Verde

Session Chair: Alberto Bacchelli

An Empirical Study of End-user Programmers in the Computer Music Community

Gregory Bulet and Abram Hindle (University of Alberta, Canada)

Are Bullies more Productive? Empirical Study of Affectiveness vs. Issue Fixing Time

Marco Ortu, Bram Adams, Giuseppe Destefanis, Parastou Tourani, Michele Marchesi and Roberto Tonelli (University of Cagliari, Italy, École polytechnique de Montréal, Canada and CRIM, The Islamic Republic of Iran)

What is the Gist? Understanding the Use of Public Gists on GitHub

Weiliang Wang, Germán Poo-Caamaño, Evan Wilde and Daniel German (University of Victoria, Canada)

Licenses, Deep Learning, and Process Mining

Sun, May 17, 17:00 – 18:00

Room: Sala Verde

Session Chair: Georgios Gousios

A Method to Detect License Inconsistencies in Large-Scale Open Source Projects

Yuhao Wu, Yuki Manabe, Tetsuya Kanda, Daniel German and Katsuro Inoue (Osaka University, Japan, Kumamoto University, Japan and University of Victoria, Canada)

Toward Deep Learning Software Repositories

Martin White, Christopher Vendome, Mario Linares-Vásquez and Denys Poshyvanyk (College of William and Mary, United States)

Identifying Software Process Management

Challenges: Survey of Practitioners in a Large Global IT Company

Monika Gupta, Ashish Sureka, Padmanabhuni Srinivas and Allahbaksh Asadullah (Indraprastha Institute of Information Technology, India and Infosys Technologies Ltd., India)



MSR 2016 (May 14-15 2016) announcement

Co-located with ICSE 2016 in Austin, Texas

<http://msrconf.org>