2018 IEEE/ACM 15th International Conference on Mining Software Repositories (MSR 2018)

Gothenburg, Sweden 27 May – 3 June 2018



IEEE Catalog Number: ISBN:

CFP1878C-POD 978-1-5386-6171-0 **Copyright © 2018, Association for Computing Machinery 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: ISBN (Print-On-Demand): ISBN (Online): ISSN: CFP1878C-POD 978-1-5386-6171-0 978-1-4503-5716-6 2574-3848

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 Web: www.proceedings.com



2018 ACM/IEEE 15th International Conference on Mining Software Repositories MSR 2018

Table of Contents

Message from ICSE 2018 General Chair	xiv
Message from MSR 2018 Chairs	xvii
Organizing Committee	xix
Program Committee - Technical Paper	xxi
Mining Challenge Program Committee	xxviii
Data Showcase Committee	xxxi
Additional Reviewers	xxxii
ICSE 2018 Sponsors and Supporters	xxxiii

MSR 2018 - Data Showcase

50K-C: A Dataset of Compilable, and Compiled, Java Projects Pedro Martins (UCI), Rohan Achar (University of California), and Cristina V. Lopes (University of California)	1
JBench: A Dataset of Data Races for Concurrency Testing Jian Gao (Tsinghua University), Xin Yang (Tsinghua University), Yu Jiang (Tsinghua University), Han Liu (Tsinghua University), Weiliang Ying (Huawei Technologies Co.), and Xian Zhang (Huawei Technologies Co.)	6
 Bugs.jar: A Large-Scale, Diverse Dataset of Real-World Java Bugs Ripon Saha (Fujitsu Labs of America), Yingjun Lyu (University of Southern California), Wing Lam (University of Illinois at Urbana-Champaign), Hiroaki Yoshida (Fujitsu Labs of America), and Mukul Prasad (Fujitsu Labs of America) 	10
A Gold Standard for Emotion Annotation in Stack Overflow Nicole Novielli (University of Bari Aldo Moro), Fabio Calefato (University of Bari Aldo Moro), and Filippo Lanubile (University of Bari Aldo Moro)	
VulinOSS: A Dataset of Security Vulnerabilities in Open-Source Systems Antonios Gkortzis (Athens University of Economics and Business), Dimitris Mitropoulos (Athens University of Economics and Business), and Diomidis Spinellis (Athens University of Economics and Business)	
A Dataset of Duplicate Pull-Requests in GitHub Yue Yu (College of Computer), Zhixing Li (College of Computer), Gang Yin (College of Computer), Tao Wang (College of Computer), and Huaimin Wang (College of Computer)	

Structured Information on State and Evolution of Dockerfiles on GitHub Gerald Schermann (University of Zurich), Sali Zumberi (University of Zurich), and Jürgen Cito (University of Zurich)	26
A Graph-Based Dataset of Commit History of Real-World Android apps Franz-Xaver Geiger (Vrije Universiteit Amsterdam), Ivano Malavolta (Vrije Universiteit Amsterdam), Luca Pascarella (Delft University of Technology), Fabio Palomba (Delft University of Technology), Dario Di Nucci (Vrije Universiteit Brussel), and Alberto Bacchelli (University of Zurich)	30
Public Git Archive: A Big Code Dataset for All	34
Word Embeddings for the Software Engineering Domain Vasiliki Efstathiou (Athens University of Economics and Business), Christos Chatzilenas (Athens University of Economics and Business), and Diomidis Spinellis (Athens University of Economics and Business)	38
npm-Miner: An Infrastructure for Measuring the Quality of the npm Registry Kyriakos Chatzidimitriou (Electrical and Computer Engineering Dept.), Michail Papamichail (Electrical and Computer Engineering Dept.), Themistoklis Diamantopoulos (Electrical and Computer Engineering Dept.), Michail Tsapanos (Electrical and Computer Engineering Dept.), and Andreas Symeonidis (Electrical and Computer Engineering Dept.)	42
CROP: Linking Code Reviews to Source Code Changes Matheus Paixao (University College London), Jens Krinke (University College London), Donggyun Han (University College London), and Mark Harman (Fcebook and University College London)	46
Developer Interaction Traces Backed by IDE Screen Recordings from Think Aloud Sessions Aiko Yamashita (OsloMet - Oslo Metropolitan University), Fabio Petrillo (Concordia University), Foutse Khomh (Polytechnique Montreal), and Yann-Gaël Guéhéneuc (Concordia University)	50
A Multi-level Dataset of Linux Kernel Patchwork Yulin Xu (School of Electronics Engineering and Computer Science) and Minghui Zhou (School of Electronics Engineering and Computer Science)	54
Documented Unix Facilities over 48 Years Diomidis Spinellis (Athens University of Economics and Business)	58

MSR 2018 - Mining Challenge

Enriched Event Streams: A General Dataset for Empirical Studies on In-IDE Activities of Software Developers	52
Sebastian Proksch (University of Zurich), Sven Amann (Technische Universität Darmstadt), and Sarah Nadi (University of Alberta)	
Comprehension Effort and Programming Activities: Related? Or Not Related?	56
The Hidden Cost of Code Completion: Understanding the Impact of the Recommendation-List Length on its Efficiency	70
Xianhao Jin (Virginia Tech) and Francisco Servant (Virginia Tech)	

Empirical Study on the Relationship Between Developer's Working Habits and Efficiency	1
Mining and Extraction of Personal Software Process Measures through IDE Interaction Logs	3
Predicting Developers' IDE Commands with Machine Learning	2
Do Software Engineers Use Autocompletion Features Differently than Other Developers?	5
 Who's This? Developer Identification Using IDE Event Data)
Detecting and Characterizing Developer Behavior Following Opportunistic Reuse of Code Snippets from the Web	4
Revisiting "Programmers' Build Errors" in the Visual Studio Context	3
Common Statement Kind Changes to Inform Automatic Program Repair	2
Studying Developer Build Issues and Debugger Usage via Timeline Analysis in Visual Studio IDE 106 Christopher Bellman (Carleton University), Ahmad Seet (Carleton University), and Olga Baysal (Carleton University)	5
Detection and Analysis of Behavioral T-Patterns in Debugging Activities)
A Study on the Use of IDE Features for Debugging	4

MSR 2018 - Technical Papers - Welcome + Keynote

Mining the Mind, Minding the Mine: Grand Challenges in Comprehension and Mining	. 118
Andrew Ko (University of Washington)	

Technical Papers - CI and Release Engineering

An Evaluation of Open-Source Software Microbenchmark Suites for Continuous Performance Assessment 1 Christoph Laaber (University of Zurich) and Philipp Leitner (Chalmers and University of Gothenburg)	19
 Studying the Impact of Adopting Continuous Integration on the Delivery Time of Pull Requests	31
 What did Really Change with the New Release of the App?	42
CLEVER: Combining Code Metrics with Clone Detection for Just-in-Time Fault Prevention and Resolution in Large Industrial Projects	53
I'm Leaving You, Travis: A Continuous Integration Breakup Story	65

Technical Papers - Modularity and Dependency

An Empirical Evaluation of OSGi Dependencies Best Practices in the Eclipse IDE	170
On the Impact of Security Vulnerabilities in the npm Package Dependency Network	181
Feature Location Using Crowd-Based Screencasts	192
Profiling Call Changes Via Motif Mining	203
Toward Predicting Architectural Significance of Implementation Issues	215

Technical Papers - Mobile

The Android Update Problem: An Empirical Study	220
Mehran Maĥmoudi (University of Alberta) and Sarah Nadi (University of Alberta)	
Why are Android Apps Removed From Google Play? A Large-Scale Empirical Study Haoyu Wang (Beijing University of Posts and Telecommunications), Hao Li (OrangeApk), Li Li (Monash University), Yao Guo (Peking University), and Guoai Xu (Beijing University of Posts and Telecommunications)	231
Anatomy of Functionality Deletion: An Exploratory Study on Mobile Apps Maleknaz Nayebi (University of Toronto), Konstantin Kuznetsov (Saarland University/CISPA), Paul Chen (University of Calgary), Andreas Zeller (Saarland University/CISPA), and Guenther Ruhe (University of Calgary)	. 243
Characterising Deprecated Android APIs Li Li (Monash University), Jun Gao (University of Luxembourg), Tegawendé Bissyandé (University of Luxembourg), Lei Ma (Harbin Institute of Technology), Xin Xia (Monash University), and Jacques Klein (University of Luxembourg)	. 254
Leveraging Historical Versions of Android Apps for Efficient and Precise Taint Analysis John Jenkins (Washington State University - Pullman) and Haipeng Cai (Washington State University - Pullman)	. 265

Technical Papers - Programming Practice

 Understanding the Usage, Impact, and Adoption of Non-OSI Approved Licenses	'0
Prevalence of Confusing Code in Software Projects: Atoms of Confusion in the Wild	;1
 How Swift Developers Handle Errors	12
 What are Your Programming Language's Energy-Delay Implications?	13

Technical Papers - 2008 Most Influential Paper Award and Evolution and Changes

What Do Large Commits Tell Us? A Taxonomical Study of Large Commits	. N/A
 SOTorrent: Reconstructing and Analyzing the Evolution of Stack Overflow Posts	19
A Design Structure Matrix Approach for Measuring Co-change-Modularity of Software Products	31
A Study on Inappropriately Partitioned Commits — How Much and What Kinds of IP Commits in Java Projects? —	36

Technical Papers - Machine Learning for SE

Data-Driven Search-Based Software Engineering 34 Vivek Nair (North Carolina State University), Amritanshu Agrawal 34 (North Carolina State University), Jianfeng Chen (North Carolina State 34 University), Wei Fu (North Carolina State University), George Mathew 36 (North Carolina State University), Tim Menzies (North Carolina State 34 University), Leandro Minku (University of Leicester), Markus Wagner 34 (The University of Adelaide), and Zhe Yu (North Carolina State 34 University) 35	41
The Open-Closed Principle of Modern Machine Learning Frameworks	53
A Benchmark Study on Sentiment Analysis for Software Engineering Research	54
 A Deep Learning Approach to Identifying Source Code in Images and Video	76
Natural Language or Not (NLoN) - A Package for Software Engineering Text Analysis Pipeline	37

Technical Papers - OSS Practices and Methods

 How Is Video Game Development Different from Software Development in Open Source?
 Which Contributions Predict Whether Developers are Accepted into GitHub Teams
 Automatic Classification of Software Artifacts in Open-Source Applications
Large-Scale Analysis of the Co-commit Patterns of the Active Developers in GitHub's Top Repositories 426 Eldan Cohen (University of Toronto) and Mariano P. Consens (University of Toronto)
 Towards Automatically Identifying Paid Open Source Developers

Technical Papers - Search and Traceability

Analyzing Requirements and Traceability Information to Improve Bug Localization Michael Rath (Technische Universität Ilmenau), David Lo (Singapore Management University), and Patrick Mäder (Technische Universität Ilmenau)	442
Towards Extracting Web API Specifications from Documentation	454
Jinqiu Yang (University of Waterioo), Erik wittern (IBM 1.J. Watson	
Research Center), Annie T.T. Ying (EquitySim), Julian Dolby (IBM T.J.	
Watson Research Center), and Lin Tan (University of Waterloo)	
Evaluating How Developers Use General-Purpose Web-Search for Code Retrieval	465
Md Masudur Rahman (University of Virginia), Jed Barson (University of	
Virginia), Sydney Paul (Clemson University), Joshua Kayani (North	
Carolina State University), Federico Andrés Lois (Codealike),	
Sebastián Fernandez Quezada (Codealike), Christopher Parnin (North	
Carolina State University), Kathryn T. Stolee (North Carolina State	
University), and Baishakhi Ray (University of Virginia)	

Learning to Mine Aligned Code and Natural Language Pairs from Stack Overflow	476
Pengcheng Yin (Carnegie Mellon University), Bowen Deng (Carnegie	
Mellon University), Edgar Chen (Carnegie Mellon University), Bogdan	
Vasilescu (Carnegie Mellon University), and Graham Neubig (Carnegie	
Mellon University)	
A Search System for Mathematical Expressions on Software Rinerias	197
A Search System for Mathematical Expressions on Software Dinaries	
Ridhi Jain (Indraprastha Institute of Information Technology), Sai	
Ridhi Jain (Indraprastha Institute of Information Technology), Sai Prathik (Indraprastha Institute of Information Technology), Venkatesh	
Ridhi Jain (Indraprastha Institute of Information Technology), Sai Prathik (Indraprastha Institute of Information Technology), Venkatesh Vinayakarao (Indraprastha Institute of Information Technology), and	
Ridhi Jain (Indraprastha Institute of Information Technology), Sai Prathik (Indraprastha Institute of Information Technology), Venkatesh Vinayakarao (Indraprastha Institute of Information Technology), and Rahul Purandare (Indraprastha Institute of Information Technology)	

Technical Papers - APIs and Code

Imprecisions Diagnostic in Source Code Deltas Guillermo de la Torre (DCC), Romain Robbes (SwSE Research Group), and Alexandre Bergel (DCC)	.492
Exploring the Use of Automated API Migrating Techniques in Practice: An Experience Report on Android Maxime Lamothe (Concordia University) and Weiyi Shang (Concordia University)	. 503
The Patch-Flow Method for Measuring Inner Source Collaboration Maximilian Capraro (Friedrich-Alexander-Universität Erlangen-Nürnberg), Michael Dorner (Friedrich-Alexander-Universität Erlangen-Nürnberg), and Dirk Riehle (Friedrich-Alexander-Universität Erlangen-Nürnberg)	. 515
Was Self-Admitted Technical Debt Removal a Real Removal? An In-Depth Perspective <i>Fiorella Zampetti (University of Sannio), Alexander Serebrenik</i> <i>(Eindhoven University of Technology), and Massimiliano Di Penta</i> <i>(University of Sannio)</i>	. 526
RestMule: Enabling Resilient Clients for Remote APIs Beatriz A. Sanchez (University of York), Konstantinos Barmpis (University of York), Patrick Neubauer (University of York), Richard F. Paige (University of York), and Dimitrios S. Kolovos (University of York)	. 537

Technical Papers - Modeling and Prediction

Deep Learning Similarities from Different Representations of Source Code	. 542
Michele Tufano (College of William and Mary), Cody Watson (College of	
William and Mary), Gabriele Bavota (Università della Svizzera italiana	
(USI)), Massimiliano Di Penta (University of Sannio), Martin White	
(College of William and Mary), and Denys Poshyvanyk (College of	
William and Mary)	
500+ Times Faster than Deep Learning: (A Case Study Exploring Faster Methods for Text Mining	
StackOverflow)	554
Tim Menzies (North Carolina State University), Suvodeep Majumder	
(North Carolina State University), Nikhila Balaji (North Carolina	
State University), Katie Brey (North Carolina State University), and	
Wei Fu (North Carolina State University)	

Studying the Relationship between Exception Handling Practices and Post-Release Defects Guilherme B. de Pádua (Concordia University) and Weiyi Shang (Concordia University)	564
Analyzing Conflict Predictors in Open-Source Java Projects Paola Accioly (Federal University of Pernambuco), Paulo Borba (Federal University of Pernambuco), Léuson Silva (Federal University of Pernambuco), and Guilherme Cavalcanti (Federal University of	576
Pernambuco)	
Bayesian Hierarchical Modelling for Tailoring Metric Thresholds Neil A. Ernst (University of Victoria)	587
Author Index	593