

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	xxxii
Additional Reviewers	xxxii
ICSE 2018 Sponsors and Supporters	xxxiii

MSR 2018 - Data Showcase

50K-C: A Dataset of Compilable, and Compiled, Java Projects	1
<i>Pedro Martins (UCI), Rohan Achar (University of California), and Cristina V. Lopes (University of California)</i>	
JBench: A Dataset of Data Races for Concurrency Testing	6
<i>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.)</i>	
Bugs.jar: A Large-Scale, Diverse Dataset of Real-World Java Bugs	10
<i>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)</i>	
A Gold Standard for Emotion Annotation in Stack Overflow	14
<i>Nicole Novielli (University of Bari Aldo Moro), Fabio Calefato (University of Bari Aldo Moro), and Filippo Lanubile (University of Bari Aldo Moro)</i>	
VulinOSS: A Dataset of Security Vulnerabilities in Open-Source Systems	18
<i>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)</i>	
A Dataset of Duplicate Pull-Requests in GitHub	22
<i>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)</i>	

Structured Information on State and Evolution of Dockerfiles on GitHub	26
<i>Gerald Schermann (University of Zurich), Sali Zumberi (University of Zurich), and Jürgen Cito (University of Zurich)</i>	
A Graph-Based Dataset of Commit History of Real-World Android apps	30
<i>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)</i>	
Public Git Archive: A Big Code Dataset for All	34
<i>Vadim Markovtsev (source{d}) and Waren Long (source{d})</i>	
Word Embeddings for the Software Engineering Domain	38
<i>Vasiliki Efsthathiou (Athens University of Economics and Business), Christos Chatzilenas (Athens University of Economics and Business), and Diomidis Spinellis (Athens University of Economics and Business)</i>	
npm-Miner: An Infrastructure for Measuring the Quality of the npm Registry	42
<i>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.)</i>	
CROP: Linking Code Reviews to Source Code Changes	46
<i>Matheus Paixao (University College London), Jens Krinke (University College London), Donggyun Han (University College London), and Mark Harman (Facebook and University College London)</i>	
Developer Interaction Traces Backed by IDE Screen Recordings from Think Aloud Sessions	50
<i>Aiko Yamashita (OsloMet - Oslo Metropolitan University), Fabio Petrillo (Concordia University), Foutse Khomh (Polytechnique Montreal), and Yann-Gaël Guéhéneuc (Concordia University)</i>	
A Multi-level Dataset of Linux Kernel Patchwork	54
<i>Yulin Xu (School of Electronics Engineering and Computer Science) and Minghui Zhou (School of Electronics Engineering and Computer Science)</i>	
Documented Unix Facilities over 48 Years	58
<i>Diomidis Spinellis (Athens University of Economics and Business)</i>	

MSR 2018 - Mining Challenge

Enriched Event Streams: A General Dataset for Empirical Studies on In-IDE Activities of Software Developers	62
<i>Sebastian Proksch (University of Zurich), Sven Amann (Technische Universität Darmstadt), and Sarah Nadi (University of Alberta)</i>	
Comprehension Effort and Programming Activities: Related? Or Not Related?	66
<i>Akond Rahman (North Carolina State University)</i>	
The Hidden Cost of Code Completion: Understanding the Impact of the Recommendation-List Length on its Efficiency	70
<i>Xianhao Jin (Virginia Tech) and Francisco Servant (Virginia Tech)</i>	

Empirical Study on the Relationship Between Developer's Working Habits and Efficiency	74
<i>Ariel Rodriguez (kyushu university), fumiya tanaka (kyushu university), and yasutaka kamei (kyushu university)</i>	
Mining and Extraction of Personal Software Process Measures through IDE Interaction Logs	78
<i>Alireza Joonbakhsh (Shiraz University) and Ashkan Sami (Shiraz University)</i>	
Predicting Developers' IDE Commands with Machine Learning	82
<i>Tyson Bulmer (University of Victoria), Lloyd Montgomery (University of Victoria), and Daniela Damian (University of Victoria)</i>	
Do Software Engineers Use Autocompletion Features Differently than Other Developers?	86
<i>Rahul Amlekar (McGill University), Andrés Felipe Rincón Gamboa (McGill University), Keheliya Gallaba (McGill University), and Shane McIntosh (McGill University)</i>	
Who's This? Developer Identification Using IDE Event Data	90
<i>John Wilkie (University College London), Ziad Al Halabi (University College London), Alperen Karaoglu (University College London), Jiafeng Liao (University College London), George Ndungu (University College London), Chaiyong Ragkhitwetsagul (University College London), Matheus Paixao (University College London), and Jens Krinke (University College London)</i>	
Detecting and Characterizing Developer Behavior Following Opportunistic Reuse of Code Snippets from the Web	94
<i>Agnieszka Ciborowska (Virginia Commonwealth University), Nicholas A. Kraft (ABB Corporate Research), and Kostadin Damevski (Virginia Commonwealth University)</i>	
Revisiting "Programmers' Build Errors" in the Visual Studio Context	98
<i>Noam Rabbani (McGill University), Michael S. Harvey (McGill University), Sadnan Saquif (McGill University), Keheliya Gallaba (McGill University), and Shane McIntosh (McGill University)</i>	
Common Statement Kind Changes to Inform Automatic Program Repair	102
<i>Mauricio Soto (Carnegie Mellon University) and Claire Le Goues (Carnegie Mellon University)</i>	
Studying Developer Build Issues and Debugger Usage via Timeline Analysis in Visual Studio IDE	106
<i>Christopher Bellman (Carleton University), Ahmad Seet (Carleton University), and Olga Baysal (Carleton University)</i>	
Detection and Analysis of Behavioral T-Patterns in Debugging Activities	110
<i>César Soto-Valero (Universidad Central "Marta Abreu" de Las Villas), Johann Bourcier (University of Rennes I-IRISA), and Benoit Baudry (KTH Royal Institute of Technology)</i>	
A Study on the Use of IDE Features for Debugging	114
<i>Afsoon Afzal (Carnegie Mellon University) and Claire Le Goues (Carnegie Mellon University)</i>	

MSR 2018 - Technical Papers - Welcome + Keynote

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

Technical Papers - CI and Release Engineering

An Evaluation of Open-Source Software Microbenchmark Suites for Continuous Performance Assessment	119
<i>Christoph Laaber (University of Zurich), Philipp Leitner (Chalmers), and University of Gothenburg undefined (undefined)</i>	
Studying the Impact of Adopting Continuous Integration on the Delivery Time of Pull Requests	131
<i>João Helis Bernardo (Federal Institute of Rio Grande do Norte), Daniel Alencar da Costa (Queen's University), and Uirá Kulesza (Federal University of Rio Grande do Norte)</i>	
What did Really Change with the New Release of the App?	142
<i>Paolo Calciati (IMDEA Software Institute / Universidad Politécnica de Madrid), Konstantin Kuznetsov (Saarland University / CISPA), Xue Bai (Beijing Institute of Technology), and Alessandra Gorla (IMDEA Software Institute)</i>	
CLEVER: Combining Code Metrics with Clone Detection for Just-in-Time Fault Prevention and Resolution in Large Industrial Projects	153
<i>Mathieu Nayrolles (La Forge Research Lab) and Abdelwahab Hamou-Lhadj (ECE Department)</i>	
I'm Leaving You, Travis: A Continuous Integration Breakup Story	165
<i>David Widder (Carnegie Mellon University), Bogdan Vasilescu (Carnegie Mellon University), Michael Hilton (Carnegie Mellon University), and Christian Kästner (Carnegie Mellon University)</i>	

Technical Papers - Modularity and Dependency

An Empirical Evaluation of OSGi Dependencies Best Practices in the Eclipse IDE	170
<i>Lina Ochoa (Centrum Wiskunde & Informatica), Thomas Degueule (Centrum Wiskunde & Informatica), and Jurgen Vinju (Centrum Wiskunde & Informatica)</i>	
On the Impact of Security Vulnerabilities in the npm Package Dependency Network	181
<i>Alexandre Decan (University of Mons), Tom Mens (University of Mons), and Eleni Constantinou (University of Mons)</i>	
Feature Location Using Crowd-Based Screencasts	192
<i>Parisa Moslehi (Concordia University), Bram Adams (Polytechnique Montreal), and Juergen Rilling (Concordia University)</i>	
Profiling Call Changes Via Motif Mining	203
<i>Barbara Russo (Free University of Bozen-Bolzano)</i>	
Toward Predicting Architectural Significance of Implementation Issues	215
<i>Arman Shahbazian (University of Southern California), Daye Nam (University of Southern California), and Nenad Medvidovic (University of Southern California)</i>	

Technical Papers - Mobile

The Android Update Problem: An Empirical Study	220
<i>Mehran Mahmoudi (University of Alberta) and Sarah Nadi (University of Alberta)</i>	
Why are Android Apps Removed From Google Play? A Large-Scale Empirical Study	231
<i>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)</i>	
Anatomy of Functionality Deletion: An Exploratory Study on Mobile Apps	243
<i>Maleknaz Nayebe (University of Toronto), Konstantin Kuznetsov (Saarland University/CISPA), Paul Chen (University of Calgary), Andreas Zeller (Saarland University/CISPA), and Guenther Ruhe (University of Calgary)</i>	
Characterising Deprecated Android APIs	254
<i>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)</i>	
Leveraging Historical Versions of Android Apps for Efficient and Precise Taint Analysis	265
<i>John Jenkins (Washington State University - Pullman) and Haipeng Cai (Washington State University - Pullman)</i>	

Technical Papers - Programming Practice

Understanding the Usage, Impact, and Adoption of Non-OSI Approved Licenses	270
<i>Rômulo Meloca (Federal University of Technology - Paraná), Gustavo Pinto (University of Pará), Leonardo Baiser (Federal University of Technology - Paraná), Marco Mattos (Federal University of Technology - Paraná), Ivanilton Polato (Federal University of Technology - Paraná), Igor Wiese (Federal University of Technology - Paraná), and Daniel M German (University of Victoria)</i>	
Prevalence of Confusing Code in Software Projects: Atoms of Confusion in the Wild	281
<i>Dan Gopstein (New York University), Henry Hongwei Zhou (New York University), Phyllis Frankl (New York University), and Justin Cappos (New York University)</i>	
How Swift Developers Handle Errors	292
<i>Nathan Cassee (Eindhoven University of Technology), Gustavo Pinto (Federal University of Pará), Fernando Castor (Federal University of Pernambuco), and Alexander Serebrenik (Eindhoven University of Technology)</i>	
What are Your Programming Language's Energy-Delay Implications?	303
<i>Stefanos Georgiou (Athens University of Economics and Business), Maria Kechagia (Delft University of Technology), Panos Louridas (Athens University of Economics and Business), and Diomidis Spinellis (Athens University of Economics and Business)</i>	

"Automatically Assessing Code Understandability" Reanalyzed: Combined Metrics Matter	314
<i>Asher Trockman (University of Evansville), Keenen Cates (University of Evansville), Mark Mozina (University of Evansville), Tuan Nguyen (University of Evansville), Christian Kästner (Carnegie Mellon University), and Bogdan Vasilescu (Carnegie Mellon University)</i>	

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

What Do Large Commits Tell Us? A Taxonomical Study of Large Commits	0
SOTorrent: Reconstructing and Analyzing the Evolution of Stack Overflow Posts	319
<i>Sebastian Baltes (University of Trier), Lorik Dumani (University of Trier), Christoph Treude (University of Adelaide), and Stephan Diehl (University of Trier)</i>	
A Design Structure Matrix Approach for Measuring Co-change-Modularity of Software Products	331
<i>Robert Benkoczi (University of Lethbridge), Daya Gaur (University of Lethbridge), Shahadat Hossain (University of Lethbridge), and Muhammad Ali Khan (University of Lethbridge)</i>	
A Study on Inappropriately Partitioned Commits — How Much and What Kinds of IP Commits in Java Projects? —	336
<i>Ryo Arima (Osaka University), Yoshiki Higo (Osaka University), and Shinji Kusumoto (Osaka University)</i>	

Technical Papers - Machine Learning for SE

Data-Driven Search-Based Software Engineering	341
<i>Vivek Nair (North Carolina State University), Amritanshu Agrawal (North Carolina State University), Jianfeng Chen (North Carolina State University), Wei Fu (North Carolina State University), George Mathew (North Carolina State University), Tim Menzies (North Carolina State University), Leandro Minku (University of Leicester), Markus Wagner (The University of Adelaide), and Zhe Yu (North Carolina State University)</i>	
The Open-Closed Principle of Modern Machine Learning Frameworks	353
<i>Houssem Ben Braiek (SWAT Lab.), Foutse Khomh (SWAT Lab.), and Bram Adams (MCIS)</i>	
A Benchmark Study on Sentiment Analysis for Software Engineering Research	364
<i>Nicole Novielli (University of Bari Aldo Moro), Daniela Girardi (University of Bari Aldo Moro), and Filippo Lanubile (University of Bari Aldo Moro)</i>	
A Deep Learning Approach to Identifying Source Code in Images and Video	376
<i>Jordan Ott (Chapman University), Abigail Atchison (Chapman University), Paul Harnack (Chapman University), Adrienne Bergh (Chapman University), and Erik Linstead (Chapman University)</i>	
Natural Language or Not (NLoN) - A Package for Software Engineering Text Analysis Pipeline	387
<i>Mika Mäntylä (University of Oulu), Fabio Calefato (University of Bari), and Maëlick Claes (University of Oulu)</i>	

Technical Papers - OSS Practices and Methods

How Is Video Game Development Different from Software Development in Open Source?	392
<i>Luca Pascarella (Delft University of Technology), Fabio Palomba (University of Zurich), Massimiliano Di Penta (University of Sannio), and Alberto Bacchelli (University of Zurich)</i>	
Which Contributions Predict Whether Developers are Accepted into GitHub Teams	403
<i>Justin Middleton (North Carolina State University), Emerson Murphy-Hill (North Carolina State University), Demetrius Green (North Carolina State University), Adam Meade (North Carolina State University), Roger Mayer (North Carolina State University), David White (North Carolina State University), and Steve McDonald (North Carolina State University)</i>	
Automatic Classification of Software Artifacts in Open-Source Applications	414
<i>Yuzhan Ma (Washington State University), Sarah Fakhoury (Washington State University), Michael Christensen (Washington State University), Venera Arnaudova (Washington State University), Waleed Zogaan (Rochester Institute of Technology), and Mehdi Mirakhorli (Rochester Institute of Technology)</i>	
Large-Scale Analysis of the Co-commit Patterns of the Active Developers in GitHub's Top Repositories.....	426
<i>Eldan Cohen (University of Toronto) and Mariano P. Consens (University of Toronto)</i>	
Towards Automatically Identifying Paid Open Source Developers	437
<i>Maëlick Claes (University of Oulu), Mika Mäntylä (University of Oulu), Miikka Kuutila (University of Oulu), and Umar Farooq (University of Oulu)</i>	

Technical Papers - Search and Traceability

Analyzing Requirements and Traceability Information to Improve Bug Localization	442
<i>Michael Rath (Technische Universität Ilmenau), David Lo (Singapore Management University), and Patrick Mäder (Technische Universität Ilmenau)</i>	
Towards Extracting Web API Specifications from Documentation	454
<i>Jinqiu Yang (University of Waterloo), Erik Wittern (IBM T.J. Watson Research Center), Annie T.T. Ying (EquitySim), Julian Dolby (IBM T.J. Watson Research Center), and Lin Tan (University of Waterloo)</i>	
Evaluating How Developers Use General-Purpose Web-Search for Code Retrieval	465
<i>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)</i>	

Learning to Mine Aligned Code and Natural Language Pairs from Stack Overflow	476
<i>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)</i>	
A Search System for Mathematical Expressions on Software Binaries	487
<i>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)</i>	

Technical Papers - APIs and Code

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

Technical Papers - Modeling and Prediction

Deep Learning Similarities from Different Representations of Source Code	542
<i>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)</i>	
500+ Times Faster than Deep Learning: (A Case Study Exploring Faster Methods for Text Mining StackOverflow)	554
<i>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)</i>	

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