

# **2022 IEEE/ACM 19th International Conference on Mining Software Repositories (MSR 2022)**

**Virtual Conference  
18 – 20 May 2022**

**Pittsburgh, Pennsylvania, USA  
23 – 24 May 2022**



**IEEE Catalog Number: CFP2278C-POD  
ISBN: 978-1-6654-5210-6**

**Copyright © 2022, Association for Computing Machinery (ACM)  
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:	CFP2278C-POD
ISBN (Print-On-Demand):	978-1-6654-5210-6
ISBN (Online):	978-1-4503-9303-4
ISSN:	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](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

CURRAN ASSOCIATES INC.  
**proceedings**  
.com

# The 2022 Mining Software Repositories Conference **MSR 2022**

## Table of Contents

Message from the MSR 2022 General and Program Co-Chairs .....	xviii
Message from the MSR 2022 Data and Tool Showcase Track Co-Chairs .....	xxiii
Message from the MSR 2022 Hackathon Track Co-Chairs .....	xxiv
Message from the MSR 2022 Industry Track Co-Chairs .....	xxv
Message from the MSR 2022 Mining Challenge Track Co-Chairs .....	xxvi
Message from the MSR 2022 Registered Reports Track Co-Chairs .....	xxviii
Message from the MSR 2022 Shadow PC Track Co-Chairs .....	xxx
Message from the MSR 2022 Tutorials Track Co-Chairs .....	xxxi
Organizing Committee .....	xxxii
Program Committee .....	xxxv

### Session 1: Communication (Documentation & Automation)

An Empirical Evaluation of GitHub Copilot's Code Suggestions .....	1
<i>Nhan Nguyen (University of Alberta, Canada) and Sarah Nadi (University of Alberta, Canada)</i>	
BotHunter: An Approach to Detect Software Bots in GitHub .....	6
<i>Ahmad Abdellatif (Concordia University, Canada), Mairieli Wessel (Delft University of Technology, Netherlands), Igor Steinmacher (Universidade Tecnológica Federal do Paraná, Brazil), Marco A. Gerosa (Northern Arizona University, USA), and Emad Shihab (Concordia University, Canada)</i>	
Comments on Comments: Where Code Review and Documentation Meet .....	18
<i>Nikitha Rao (Carnegie Mellon University, United States), Jason Tsay (IBM Research, United States), Martin Hirzel (IBM Research, United States), and Vincent J. Hellendoorn (Carnegie Mellon University, United States)</i>	
Does This Apply to Me? An Empirical Study of Technical Context in Stack Overflow .....	23
<i>Akalanka Galappaththi (University of Alberta, Canada), Sarah Nadi (University of Alberta, Canada), and Christoph Treude (University of Melbourne, Australia)</i>	
Towards Reliable Agile Iterative Planning via Predicting Documentation Changes of Work Items .....	35
<i>Jirat Pasuksmit (The University of Melbourne, Australia), Patanamon Thongtanunam (The University of Melbourne, Australia), and Shanika Karunasekera (The University of Melbourne, Australia)</i>	

## Session 2: Maintenance (Issues & Smells)

Beyond Duplicates: Towards Understanding and Predicting Link Types in Issue Tracking Systems .....	48
<i>Clara Marie Lüders (University of Hamburg, Germany), Abir Bouraffa (University of Hamburg, Germany), and Walid Maalej (University of Hamburg, Germany)</i>	
Smelly Variables in Ansible Infrastructure Code: Detection, Prevalence, and Lifetime .....	61
<i>Ruben Opdebeeck (Vrije Universiteit Brussel, Belgium), Ahmed Zerouali (Vrije Universiteit Brussel, Belgium), and Coen De Roover (Vrije Universiteit Brussel, Belgium)</i>	
An Alternative Issue Tracking Dataset of Public Jira Repositories .....	73
<i>Lloyd Montgomery (University of Hamburg, Germany), Clara Lüders (University of Hamburg, Germany), and Walid Maalej (University of Hamburg, Germany)</i>	
Real-World Clone-Detection in Go .....	78
<i>Qinyun Wu (Bytedance Ltd., China), Huan Song (Bytedance Ltd., China), and Ping Yang (Bytedance Ltd., China)</i>	

## Session 3: Introspection, Vision, and Human Aspects

Geographic Diversity in Public Code Contributions: An Exploratory Large-Scale Study Over 50 Years .....	80
<i>Davide Rossi (University of Bologna, Italy) and Stefano Zacchiroli (LTCI, Télécom Paris, Institut Polytechnique de Paris, France)</i>	
Operationalizing Threats to MSR Studies by Simulation-Based Testing .....	86
<i>Johannes Härtel (University of Koblenz, Germany) and Ralf Lämmel (University of Koblenz, Germany)</i>	
The General Index of Software Engineering Papers .....	98
<i>Zeinab Abou Khalil (Inria de Paris, France) and Stefano Zacchiroli (LTCI, Télécom Paris, Institut Polytechnique de Paris, France)</i>	
Challenges and Future Research Direction for Microtask Programming in Industry .....	103
<i>Masanari Kondo (Kyushu University, Japan), Shinobu Saito (NTT Computer and Data Science Laboratories, Japan), Yukako Imura (NTT Computer and Data Science Laboratories, Japan), Eunjong Choi (Kyoto Institute of Technology, Japan), Osamu Mizuno (Kyoto Institute of Technology, Japan), Yasutaka Kamei (Kyushu University, Japan), and Naoyasu Ubayashi (Kyushu University, Japan)</i>	
Starting the InnerSource Journey: Key Goals and Metrics to Measure Collaboration .....	105
<i>Daniel Izquierdo-Cortázar (Bitergia, Spain), Jesús Alonso-Gutiérrez (Santander Group, Spain), Alberto Pérez García-Plaza (Bitergia, Spain), Gregorio Robles (Universidad Rey Juan Carlos, Spain), and Jesús M. González-Barahona (Universidad Rey Juan Carlos, Spain)</i>	

## Mining Challenge

An Exploratory Study on Refactoring Documentation in Issues Handling .....	107
<i>Eman Abdullah AlOmar (Stevens Institute of Technology, USA), Anthony Peruma (Rochester Institute of Technology, USA), Mohamed Wiem Mkaouer (Rochester Institute of Technology, USA), Christian D. Newman (Rochester Institute of Technology, USA), and Ali Ouni (University of Quebec, Canada)</i>	
Between JIRA and GitHub: ASFBot and its Influence on Human Comments in Issue Trackers .....	112
<i>Ambarish Moharil (Eindhoven University of Technology, The Netherlands), Dmitrii Orlov (Eindhoven University of Technology, The Netherlands), Samar Jameel (Eindhoven University of Technology, The Netherlands), Tristan Trouwen (Eindhoven University of Technology, The Netherlands), Nathan Cassee (Eindhoven University of Technology, The Netherlands), and Alexander Serebrenik (Eindhoven University of Technology, The Netherlands)</i>	
Is Refactoring Always a Good Egg? Exploring the Interconnection Between Bugs and Refactorings .....	117
<i>Amirreza Bagheri (University of Szeged, Hungary) and Peter Hegedus (University of Szeged, Hungary)</i>	
On the Co-Occurrence of Refactoring of Test and Source Code .....	122
<i>Nicholas Alexandre Nagy (Concordia University, Canada) and Rabe Abdalkareem (Carleton University, Canada)</i>	
Refactoring Debt: Myth or Reality? An Exploratory Study on the Relationship Between Technical Debt and Refactoring .....	127
<i>Anthony Peruma (Rochester Institute of Technology, USA), Eman Abdullah AlOmar (Stevens Institute of Technology, USA), Christian D. Newman (Rochester Institute of Technology, USA), Mohamed Wiem Mkaouer (Rochester Institute of Technology, USA), and Ali Ouni (University of Quebec, Canada)</i>	
Studying the Impact of Continuous Delivery Adoption on Bug-Fixing Time in Apache's Open-Source Projects .....	132
<i>Carlos Diego Andrade de Almeida (Federal University of Ceará, Brazil), Diego Nogueira Feijó (Federal University of Ceará, Brazil), and Lincoln Souza Rocha (Federal University of Ceará, Brazil)</i>	
Which Bugs are Missed in Code Reviews: An Empirical Study on SmartSHARK Dataset .....	137
<i>Fatemeh Khoshnoud (Shiraz University, Iran), Ali Rezaei Nasab (Shiraz University, Iran), Zahra Toudeji (Shiraz University, Iran), and Ashkan Sami (Shiraz University, Iran)</i>	

## Tutorial

Empirical Standards for Repository Mining .....	142
<i>Preetha Chatterjee (Drexel University, USA), Tushar Sharma (Dalhousie University, Canada), and Paul Ralph (Dalhousie University, Canada)</i>	

## Session 4: Software Quality (Bugs & Smells)

Dazzle: using Optimized Generative Adversarial Networks to Address Security Data Class Imbalance Issue .....	144
<i>Rui Shu (North Carolina State University, USA), Tianpei Xia (North Carolina State University, USA), Laurie Williams (North Carolina State University, USA), and Tim Menzies (North Carolina State University, USA)</i>	
How to Improve Deep Learning for Software Analytics (a Case Study with Code Smell Detection) .....	156
<i>Rahul Yedida (NC State University, USA) and Tim Menzies (NC State University, USA)</i>	
To What Extent do Deep Learning-Based Code Recommenders Generate Predictions by Cloning Code from the Training Set? .....	167
<i>Matteo Ciniselli (Università della Svizzera italiana, Switzerland), Luca Pascarella (Università della Svizzera italiana, Switzerland), and Gabriele Bavota (Università della Svizzera italiana, Switzerland)</i>	
Searching for High-Fidelity Builds using Active Learning .....	179
<i>Harshitha Menon (Lawrence Livermore National Laboratory, USA), Konstantinos Parasyris (Lawrence Livermore National Laboratory, USA), Tom Scogland (Lawrence Livermore National Laboratory, USA), and Todd Gamblin (Lawrence Livermore National Laboratory, USA)</i>	
ApacheJIT: A Large Dataset for Just-In-Time Defect Prediction .....	191
<i>Hossein Keshavarz (University of Waterloo, Canada) and Meiyappan Nagappan (University of Waterloo, Canada)</i>	
ReCover: a Curated Dataset for Regression Testing Research .....	196
<i>Francesco Altiero (Università degli Studi di Napoli Federico II, Italy), Anna Corazza (Università degli Studi di Napoli Federico II, Italy), Sergio Di Martino (Università degli Studi di Napoli Federico II, Italy), Adriano Peron (Università degli Studi di Napoli Federico II, Italy), and Luigi L. L. Starace (Università degli Studi di Napoli Federico II, Italy)</i>	

## Tutorial

Mining the Ethereum Blockchain Platform: Best Practices and Pitfalls (MSR 2022 Tutorial) .....	201
<i>Gustavo A. Oliva (Queen's University, Canada)</i>	

## Session 5: Communication & Domains

Mining the Usage of Reactive Programming APIs: A Study on GitHub and Stack Overflow .....	203
<i>Carlos Zimmerle (Federal University of Pernambuco, Brazil), Kiev Gama (Federal University of Pernambuco, Brazil), Fernando Castor (Utrecht University, The Netherlands), and José Murilo Mota Filho (Federal University of Pernambuco, Brazil)</i>	

Painting the Landscape of Automotive Software in GitHub .....	215
<i>Sangeeth Kochanthara (Eindhoven University of Technology, The Netherlands), Yanja Dajsuren (Eindhoven University of Technology, The Netherlands), Loek Cleophas (Eindhoven University of Technology, The Netherlands), and Mark van den Brand (Eindhoven University of Technology, The Netherlands)</i>	
DISCO: A Dataset of Discord Chat Conversations for Software Engineering Research .....	227
<i>Keerthana Muthu Subash (Carleton University, Canada), Lakshmi Prasanna Kumar (Carleton University, Canada), Sri Lakshmi Vadlamani (Carleton University, Canada), Preetha Chatterjee (Drexel University, United States), and Olga Baysal (Carleton University, Canada)</i>	
Inspect4py: A Knowledge Extraction Framework for Python Code Repositories .....	232
<i>Rosa Filgueira (University of St Andrews, UK) and Daniel Garijo (Universidad Politécnica de Madrid, Spain)</i>	
SLNET: A Redistributable Corpus of 3rd-Party Simulink Models .....	237
<i>Sohil Lal Shrestha (University of Texas at Arlington, USA), Shafiul Azam Chowdhury (University of Texas at Arlington, USA), and Christoph Csallner (University of Texas at Arlington, USA)</i>	
SoCCMiner: A Source Code-Comments and Comment-Context Miner .....	242
<i>Murali Sridharan (University of Oulu, Finland), Mika Mäntylä (University of Oulu, Finland), Maëlick Claes (University of Oulu, Finland), and Leevi Rantala (University of Oulu, Finland)</i>	
SOSum: A Dataset of Stack Overflow Post Summaries .....	247
<i>Bonan Kou (Purdue University, USA), Yifeng Di (Purdue University, USA), Muhao Chen (University of Southern California, USA), and Tianyi Zhang (Purdue University, USA)</i>	

## Session 6: Maintenance & Testing

An Empirical Study on Maintainable Method Size in Java .....	252
<i>Shaiful Alam Chowdhury (University of British Columbia, Canada), Gias Uddin (University of Calgary, Canada), and Reid Holmes (University of British Columbia, Canada)</i>	
Characterizing High-Quality Test Methods: A First Empirical Study .....	265
<i>Victor Veloso (Universidade Federal de Minas Gerais, Brazil) and Andre Hora (Universidade Federal de Minas Gerais, Brazil)</i>	
CLIP Meets GamePhysics: Towards bug Identification in Gameplay Videos using Zero-Shot Transfer Learning .....	270
<i>Mohammad Reza Taesiri (University of Alberta, Canada), Finlay Macklon (University of Alberta, Canada), and Cor-Paul Bezemer (University of Alberta, Canada)</i>	
Complex Python Features in the Wild .....	282
<i>Yi Yang (Rensselaer Polytechnic Institute, USA), Ana Milanova (Rensselaer Polytechnic Institute, USA), and Martin Hirzel (IBM Research, USA)</i>	

ManyTypes4TypeScript: A Comprehensive TypeScript Dataset for Sequence-Based Type Inference.....	294
<i>Kevin Jesse (University of California, Davis, USA) and Premkumar T. Devanbu (University of California, Davis, USA)</i>	
Methods2Test: A Dataset of Focal Methods Mapped to Test Cases .....	299
<i>Michele Tufano (Microsoft, USA), Shao Kun Deng (Microsoft, USA), Neel Sundaresan (Microsoft, USA), and Alexey Svyatkovskiy (Microsoft, USA)</i>	
npm-Filter: Automating the Mining of Dynamic Information from npm Packages .....	304
<i>Ellen Arteca (Northeastern University, USA) and Alexi Turcotte (Northeastern University, USA)</i>	

## Session 7: Developer Wellbeing & Project Communication

How Heated is it? Understanding GitHub Locked Issues .....	309
<i>Isabella Ferreira (Polytechnique Montréal, Canada), Bram Adams (Queen's University, Canada), and Jinghui Cheng (Polytechnique Montréal, Canada)</i>	
On the Violation of Honesty in Mobile Apps: Automated Detection and Categories .....	321
<i>Humphrey O. Obie (Monash University, Australia), Idowu Ileku (Data Science Nigeria, Nigeria), Hung Du (Deakin University, Australia), Mojtaba Shahin (RMIT University, Australia), John Grundy (Monash University, Australia), Li Li (Monash University, Australia), Jon Whittle (CSIRO's Data61, Australia), and Burak Turhan (University of Oulu, Finland)</i>	
Exploring Apache Incubator Project Trajectories with APEX .....	333
<i>Anirudh Ramchandran (University of California, Davis, USA), Likang Yin (University of California, Davis, USA), and Vladimir Filkov (University of California, Davis, USA)</i>	
The OCEAN Mailing List Data Set: Network Analysis Spanning Mailing Lists and Code Repositories .....	338
<i>Melanie Warrick (Google, Inc.), Samuel F. Rosenblatt (University of Vermont), Jean-Gabriel Young (University of Vermont), Amanda Casari (Google, Inc.), Laurent Hébert-Dufresne (University of Vermont), and James Bagrow (University of Vermont)</i>	
The Unexplored Treasure Trove of Phabricator Code Reviews .....	343
<i>Gunnar Kudrjavets (University of Groningen, Netherlands), Nachiappan Nagappan (Microsoft Research, USA), and Ayushi Rastogi (University of Groningen, Netherlands)</i>	
The Unsolvable Problem or the Unheard Answer? A Dataset of 24,669 Open-Source Software Conference Talks .....	348
<i>Kimberly Truong (Oregon State University, USA), Courtney Miller (Carnegie Mellon University, USA), Bogdan Vasilescu (Carnegie Mellon University, USA), and Christian Kästner (Carnegie Mellon University, USA)</i>	



## Session 8: Large-Scale Mining & Software Ecosystems

A Large-Scale Comparison of Python Code in Jupyter Notebooks and Scripts .....	353
<i>Konstantin Grotov (JetBrains Research, ITMO University), Sergey Titov (JetBrains Research), Vladimir Sotnikov (JetBrains Research), Yaroslav Golubev (JetBrains Research), and Timofey Bryksin (JetBrains Research)</i>	
An Empirical Study on the Survival Rate of GitHub Projects .....	365
<i>Adem Ait (IN3 - UOC, Spain), Javier Luis Cánovas Izquierdo (IN3 - UOC, Spain), and Jordi Cabot (IN3 - UOC, ICREA, Spain)</i>	
Do Customized Android Frameworks Keep Pace with Android? .....	376
<i>Pei Liu (Monash University, Australia), Mattia Fazzini (University of Minnesota, United States), John Grundy (Monash University, Australia), and Li Li (Monash University, Australia)</i>	
DaSEA – A Dataset for Software Ecosystem Analysis .....	388
<i>Petya Buchkova (IT University of Copenhagen, Denmark), Joakim Hey Hinnerskov (IT University of Copenhagen, Denmark), Kasper Olsen (IT University of Copenhagen, Denmark), and Rolf-Helge Pfeiffer (IT University of Copenhagen, Denmark)</i>	
Dataset: Dependency Networks of Open Source Libraries Available Through CocoaPods, Carthage and Swift PM .....	393
<i>Kristiina Rahkema (University of Tartu, Estonia) and Dietmar Pfahl (University of Tartu, Estonia)</i>	
Lupa: A Framework for Large Scale Analysis of the Programming Language Usage .....	398
<i>Anna Vlasova (JetBrains Research), Maria Tigina (JetBrains Research, ITMO University), Ilya Vlasov (Saint Petersburg State University), Anastasiia Birillo (JetBrains Research), Yaroslav Golubev (JetBrains Research), and Timofey Bryksin (JetBrains Research)</i>	
GitDelver Enterprise Dataset (GDED): An Industrial Closed-Source Dataset for Socio-Technical Research .....	403
<i>Nicolas Riquet (University of Namur, Belgium), Xavier Devroey (University of Namur, Belgium), and Benoît Vanderose (University of Namur, Belgium)</i>	

## Session 9: Scaling & Cloud

SniP: An Efficient Stack Tracing Framework for Multi-Threaded Programs .....	408
<i>Arun KP (Indian Institute of Technology, India), Saurabh Kumar (Indian Institute of Technology, India), Debadatta Mishra (Indian Institute of Technology, India), and Biswabandan Panda (Indian Institute of Technology, India)</i>	
Tooling for Time- and Space-Efficient git Repository Mining .....	413
<i>Fabian Heseding (University of Potsdam, Germany), Willy Scheibel (University of Potsdam, Germany), and Jürgen Döllner (University of Potsdam, Germany)</i>	
TSSB-3M: Mining Single Statement Bugs at Massive Scale .....	418
<i>Cedric Richter (University of Oldenburg, Germany) and Heike Wehrheim (University of Oldenburg, Germany)</i>	

## Session 10: Security

LibDB: An Effective and Efficient Framework for Detecting Third-Party Libraries in Binaries .....	423
<i>Wei Tang (Tsinghua University, China), Yanlin Wang (Microsoft Research, China), Hongyu Zhang (The University of Newcastle, Australia), Shi Han (Microsoft Research, China), Ping Luo (Tsinghua University, China), and Dongmei Zhang (Microsoft Research, China)</i>	
Noisy Label Learning for Security Defects .....	435
<i>Roland Croft (CREST - The Centre for Research on Engineering Software Technologies, University of Adelaide, Cyber Security Cooperative Research Centre, Australia), M. Ali Babar (CREST - The Centre for Research on Engineering Software Technologies, University of Adelaide, Cyber Security Cooperative Research Centre, Australia), and Huaming Chen (CREST - The Centre for Research on Engineering Software Technologies, University of Adelaide, Cyber Security Cooperative Research Centre, Australia)</i>	
WeakSATD: Detecting Weak Self-Admitted Technical Debt .....	448
<i>Barbara Russo (Free University of Bozen-Bolzano, Italy), Matteo Camilli (Free University of Bozen-Bolzano, Italy), and Moritz Mock (Free University of Bozen-Bolzano, Italy)</i>	
AndroOBFS: Time-Tagged Obfuscated Android Malware Dataset with Family Informationn .....	454
<i>Saurabh Kumar (Indian Institute of Technology Kanpur, India), Debadatta Mishra (Indian Institute of Technology Kanpur, India), Biswabandan Panda (Indian Institute of Technology Bombay, India), and Sandeep Kumar Shukla (Indian Institute of Technology Kanpur, India)</i>	
TriggerZoo: A Dataset of Android Applications Automatically Infected with Logic Bombs .....	459
<i>Jordan Samhi (University of Luxembourg, Luxembourg), Tegawendé F. Bissyandé (University of Luxembourg, Luxembourg), and Jacques Klein (University of Luxembourg, Luxembourg)</i>	
Vul4J: A Dataset of Reproducible Java Vulnerabilities Geared Towards the Study of Program Repair Techniques .....	464
<i>Quang-Cuong Bui (Hamburg University of Technology, Germany), Riccardo Scandariato (Hamburg University of Technology, Germany), and Nicolás E. Díaz Ferreyra (Hamburg University of Technology, Germany)</i>	

## Session 11: Machine Learning & Information Retrieval

Challenges in Migrating Imperative Deep Learning Programs to Graph Execution: An Empirical Study .....	469
<i>Tatiana Castro Vélez (City University of New York (CUNY) Graduate Center, USA), Raffi Khatchadourian (City University of New York (CUNY) Hunter College, USA), Mehdi Bagherzadeh (Oakland University, USA), and Anita Raja (City University of New York (CUNY) Hunter College, USA)</i>	
Does Configuration Encoding Matter in Learning Software Performance? An Empirical Study on Encoding Schemes .....	482
<i>Jingzhi Gong (Loughborough University, UK) and Tao Chen (Loughborough University, UK)</i>	

Multimodal Recommendation of Messenger Channels .....	495
<i>Ekaterina Koshchenko (JetBrains Research, The Netherlands), Egor Klimov (JetBrains Research, Russia), and Vladimir Kovalenko (JetBrains Research, The Netherlands)</i>	
On the Naturalness of Fuzzer-Generated Code .....	506
<i>Rajeswari Hita Kambhamettu (Carnegie Mellon University, USA), John Billos (Wake Forest University, USA), Tomi Oluwaseun-Apo (Pennsylvania State University, USA), Benjamin Gafford (Carnegie Mellon University, USA), Rohan Padhye (Carnegie Mellon University, USA), and Vincent J. Hellendoorn (Carnegie Mellon University, USA)</i>	
Senatus - A Fast and Accurate Code-to-Code Recommendation Engine .....	511
<i>Fran Silavong (JPMorgan Chase, United Kingdom), Sean Moran (JPMorgan Chase, United Kingdom), Antonios Georgiadis (JPMorgan Chase, United Kingdom), Rohan Saphal (JPMorgan Chase, United Kingdom), and Rob Otter (JPMorgan Chase, United Kingdom)</i>	
GraphCode2Vec: Generic Code Embedding via Lexical and Program Dependence Analyses .....	524
<i>Wei Ma (University of Luxembourg), Mengjie Zhao (LMU Munich, Germany), Ezekiel Soremekun (University of Luxembourg), Qiang Hu (University of Luxembourg), Jie M. Zhang (University College London, United Kingdom), Mike Papadakis (University of Luxembourg), Maxime Cordy (University of Luxembourg), Xiaofei Xie (Singapore Management University, Singapore), and Yves Le Traon (University of Luxembourg)</i>	

## Session 12: Integration & Large-Scale Mining

Do Small Code Changes Merge Faster? A Multi-Language Empirical Investigation .....	537
<i>Gunnar Kudrjavets (University of Groningen, Netherlands), Nachiappan Nagappan (Meta Platforms, Inc., USA), and Ayushi Rastogi (University of Groningen, Netherlands)</i>	
FaST: A Linear Time Stack Trace Alignment Heuristic for Crash Report Deduplication .....	549
<i>Irving Muller Rodrigues (Polytechnique Montreal, Canada), Daniel Aloise (Polytechnique Montreal, Canada), and Eraldo Rezende Fernandes (Leuphana University of Lüneburg, Germany)</i>	
Is Open Source Eating the World's Software? Measuring the Proportion of Open Source in Proprietary Software using Java Binaries .....	561
<i>Julius Musseau (Mergebase, Canada), John Speed Meyers (Chainguard, USA), George P. Sieniawski (IQT Labs, USA), C. Albert Thompson (Ford Motor Company, USA), and Daniel M. German (University of Victoria, Canada)</i>	
Methods for Stabilizing Models Across Large Samples of Projects (with Case Studies on Predicting Defect and Project Health) .....	566
<i>Suvodeep Majumder (North Carolina State University, USA), Tianpei Xia (North Carolina State University, USA), Rahul Krishna (North Carolina State University, USA), and Tim Menzies (North Carolina State University, USA)</i>	

Mining Code Review Data to Understand Waiting Times Between Acceptance and Merging: An Empirical Analysis .....	579
<i>Gunnar Kudrjavets (University of Groningen, Netherlands), Aditya Kumar (Snap, Inc., USA), Nachiappan Nagappan (Meta Platforms, Inc., USA), and Ayushi Rastogi (University of Groningen, Netherlands)</i>	
TwinDroid: A Dataset of Android app System Call Traces and Trace Generation Pipeline .....	591
<i>Asma Razgallah (Univeristé du Québec à Chicoutimi, Canada), Raphaël Khoury (Univeristé du Québec à Chicoutimi, Canada), and Jean-Baptiste Poulet (Univeristé du Québec à Chicoutimi, Canada)</i>	

## Session 13: Security & Quality

LineVD: Statement-Level Vulnerability Detection using Graph Neural Networks .....	596
<i>David Hin (The University of Adelaide, Australia), Andrey Kan (AWS AI Labs, Australia), Huaming Chen (The University of Adelaide, Australia), and M. Ali Babar (The University of Adelaide, Australia)</i>	
LineVul: A Transformer-Based Line-Level Vulnerability Prediction .....	608
<i>Michael Fu (Monash University, Australia) and Chakkrit Tantithamthavorn (Monash University, Australia)</i>	
On the Use of Fine-Grained Vulnerable Code Statements for Software Vulnerability Assessment Models .....	621
<i>Triet Huynh Minh Le (The University of Adelaide, Australia) and M. Ali Babar (The University of Adelaide, Australia; Cyber Security Cooperative Research Centre, Australia)</i>	
ECench: An Energy Bug Benchmark of Ethereum Client Software .....	634
<i>Jinyoung Kim (Sungkyunkwan University, Republic of Korea), Misoo Kim (Sungkyunkwan University, Republic of Korea), and Eunseok Lee (Sungkyunkwan University, Republic of Korea)</i>	
Microsoft CloudMine: Data Mining for the Executive Order on Improving the Nation's Cybersecurity .....	639
<i>Kim Herzig (Microsoft Corporation, USA), Luke Ghostling (Microsoft Corporation, USA), Maximilian Grothusmann (Microsoft Corporation, Germany), Sascha Just (Microsoft Corporation, Germany), Nora Huang (Microsoft Corporation, Canada), Alan Klimowski (Microsoft Corporation, USA), Yashasvini Ramkumar (Microsoft Corporation, USA), Myles McLeroy (Microsoft Corporation, USA), Kivanc Muslu (Microsoft Corporation, USA), Hitesh Sajani (Microsoft Corporation, USA), and Varsha Vadaga (Microsoft Corporation, USA)</i>	

## Session 14: Software Quality

Evaluating the Effectiveness of Local Explanation Methods on Source Code-Based Defect Prediction Models .....	640
<i>Gao Yuxiang (Jiangsu Normal University, China), Zhu Yi (Jiangsu Normal University, China), and Yu Qiao (Jiangsu Normal University, China)</i>	

Problems and Solutions in Applying Continuous Integration and Delivery to 20 Open-Source Cyber-Physical Systems .....	646
<i>Fiorella Zampetti (University of Sannio, Italy), Vittoria Nardone (University of Sannio, Italy), and Massimiliano Di Penta (University of Sannio, Italy)</i>	
To Type or Not to Type? A Systematic Comparison of the Software Quality of JavaScript and TypeScript Applications on GitHub .....	658
<i>Justus Bogner (University of Stuttgart, Germany) and Manuel Merkel (University of Stuttgart, Germany)</i>	
Using Bandit Algorithms for Selecting Feature Reduction Techniques in Software Defect Prediction .....	670
<i>Masateru Tsunoda (Kindai University, Japan), Akito Monden (Okayama University, Japan), Koji Toda (Fukuoka Institute of Technology, Japan), Amjed Tahir (Massey University, New Zealand), Kwabena Ebo Bennin (Wageningen University &amp; Research, Netherlands), Keitaro Nakasai (National Institute of Technology, Kagoshima College, Japan), Masataka Nagura (Nanzan University, Japan), and Kenichi Matsumoto (Nara Institute of Science and Technology, Japan)</i>	
Constructing Dataset of Functionally Equivalent Java Methods using Automated Test Generation Techniques .....	682
<i>Yoshiki Higo (Osaka University, Japan), Shinsuke Matsumoto (Osaka University, Japan), Shinji Kusumoto (Osaka University, Japan), and Kazuuya Yasuda (Hitachi, Ltd., Japan)</i>	
Extracting Corrective Actions from Code Repositories .....	687
<i>Yegor Bugayenko (Huawei, Russia), Kirill Daniakin (Innopolis University, Russia), Mirko Farina (Innopolis University, Russia), Firas Jolha (Innopolis University, Russia), Artem Kruglov (Innopolis University, Russia), Giancarlo Succi (Innopolis University, Russia), and Witold Pedrycz (University of Alberta, Canada)</i>	

## Session 15: Collaboration & Open Source

Code Review Practices for Refactoring Changes: An Empirical Study on OpenStack .....	689
<i>Eman Abdullah AlOmar (Stevens Institute of Technology, USA), Moataz Chouchen (University of Quebec, Canada), Mohamed Wiem Mkaouer (Rochester Institute of Technology, USA), and Ali Ouni (University of Quebec, Canada)</i>	
A Time Series-Based Dataset of Open-Source Software Evolution .....	702
<i>Bruno L. Sousa (Department of Computer Science - UFMG, Brazil), Mariza A. S. Bigonha (Department of Computer Science - UFMG, Brazil), Kecia A. M. Ferreira (Department of Computing - CEFET-MG, Brazil), and Glaura Franco (Department of Statistics - UFMG, Brazil)</i>	
A Versatile Dataset of Agile Open Source Software Projects .....	707
<i>Vali Tawosi (University College London, UK), Afnan Al-Subaihin (University College London, UK), Rebecca Moussa (University College London, UK), and Federica Sarro (University College London, UK)</i>	

FixJS: A Dataset of Bug-Fixing JavaScript Commits .....	712
<i>Viktor Csuwik (University of Szeged, Hungary) and László Vidács (University of Szeged, Hungary)</i>	
LAGOON: An Analysis Tool for Open Source Communities .....	717
<i>Sourya Dey (Galois, Inc., USA) and Walt Woods (Galois, Inc., USA)</i>	
Automatically Prioritizing and Assigning Tasks from Code Repositories in Puzzle Driven Development .....	722
<i>Yegor Bugayenko (Huawei, Russia), Ayomide Bakare (Innopolis University, Russia), Arina Cheverda (Innopolis University, Russia), Mirko Farina (Innopolis University, Russia), Artem Kruglov (Innopolis University, Russia), Yaroslav Plaksin (Innopolis University, Russia), Giancarlo Succi (Innopolis University, Russia), and Witold Pedrycz (University of Alberta, Canada)</i>	

## Tutorial

Software Bots in Software Engineering: Benefits and Challenges .....	724
<i>Mairieli Wessel (Delft University of Technology, The Netherlands), Marco A. Gerosa (Northern Arizona University, USA), and Emad Shihab (Concordia University, Canada)</i>	

## Hackathon

Bot Detection in GitHub Repositories .....	726
<i>Natarajan Chidambaram (University of Mons, Belgium) and Pooya Rostami Mazrae (University of Mons, Belgium)</i>	
GitRank: A Framework to Rank GitHub Repositories .....	729
<i>Niranjan Hasabnis (Intel Lab, USA)</i>	
Maintenance and Evolution: GrimoireLab Graal .....	732
<i>Willem Meijer (University of Groningen, The Netherlands), David Visscher (University of Groningen, The Netherlands), Erwin de Haan (University of Groningen, The Netherlands), Merijn Schröder (University of Groningen, The Netherlands), Leon Visscher (University of Groningen, The Netherlands), Andrea Capiluppi (University of Groningen, The Netherlands), and Ioan Botez (University of Groningen, The Netherlands)</i>	
OpenSSL 3.0.0: An Exploratory Case Study .....	735
<i>James Walden (Northern Kentucky University, USA)</i>	
Quid Pro Quo: An Exploration of Reciprocity in Code Review .....	738
<i>Gavidia-Calderon Carlos (The Open University, United Kingdom), Han DongGyun (Singapore Management University, Singapore), and Bennaceur Amel (The Open University, United Kingdom)</i>	
Replicating Data Pipelines with GrimoireLab .....	741
<i>Kalvin Eng (University of Alberta, Canada) and Hareem Sahar (University of Alberta, Canada)</i>	

## Session 16: Non-functional Properties (Availability, Security, Legal Aspects)

A Deep Study of the Effects and Fixes of Server-Side Request Races in Web Applications .....	744
<i>Zhengyi Qiu (North Carolina State University, USA), Shudi Shao (North Carolina State University, USA), Qi Zhao (North Carolina State University, USA), Hassan Ali Khan (North Carolina State University, USA), Xinning Hui (North Carolina State University, USA), and Guoliang Jin (North Carolina State University, USA)</i>	
A Large-Scale Dataset of (Open Source) License Text Variants .....	757
<i>Stefano Zacchiroli (LTCI, Télécom Paris, Institut Polytechnique de Paris, France)</i>	
Detecting Privacy-Sensitive Code Changes with Language Modeling .....	762
<i>Gökalp Demirci (Meta Platforms, Inc., USA), Vijayaraghavan Murali (Meta Platforms, Inc., USA), Imad Ahmad (Meta Platforms, Inc., USA), Rajeev Rao (Meta Platforms, Inc., USA), and Gareth Ari Aye (Meta Platforms, Inc., USA)</i>	
SECOM: Towards a Convention for Security Commit Messages .....	764
<i>Sofia Reis (University of Lisbon, Portugal), Rui Abreu (University of Porto, Portugal), Hakan Erdogmus (Carnegie Mellon University, USA), and Corina Păsăreanu (Carnegie Mellon University, USA)</i>	
Varangian: A Git Bot for Augmented Static Analysis .....	766
<i>Saurabh Pujar (IBM Research, USA), Yunhui Zheng (IBM Research, USA), Luca Buratti (IBM Research, USA), Burn Lewis (IBM Research, USA), Alessandro Morari (IBM Research, USA), Jim Laredo (IBM Research, USA), Kevin Postlethwait (Red Hat, USA), and Christoph Görn (Red Hat, Germany)</i>	
<b>Author Index</b> .....	<b>769</b>