

**2019 IEEE/ACM 16th  
International Conference on  
Mining Software Repositories  
(MSR 2019)**

**Montreal, Quebec, Canada  
25-31 May 2019**



**IEEE Catalog Number: CFP1978C-POD  
ISBN: 978-1-7281-3370-6**

**Copyright © 2019 by the Institute of Electrical and Electronics Engineers, Inc.  
All Rights Reserved**

*Copyright and Reprint Permissions:* Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. 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:	CFP1978C-POD
ISBN (Print-On-Demand):	978-1-7281-3370-6
ISBN (Online):	978-1-7281-3412-3
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

# 2019 IEEE/ACM 16th International Conference on Mining Software Repositories (MSR) **MSR 2019**

## Table of Contents

Message from the Chairs of MSR 2019	xiv
Organizing Committee for MSR 2019	xvii
Program Committee	xix
Steering Committee	xxx
Additional Reviewers	xxxiii
Keynote Abstract	xxxiv

### Representations for Mining (Part 1)

SCOR: Source Code Retrieval with Semantics and Order	1
<i>Shayan Akbar (Purdue University) and Avinash Kak (Purdue University)</i>	
PathMiner: A Library for Mining of Path-Based Representations of Code	13
<i>Vladimir Kovalenko (TU Delft), Egor Bogomolov (JetBrains Research and Higher School of Economics, Russia), Timofey Bryksin (JetBrains Research and Saint Petersburg State University, Russia), and Alberto Bacchelli (University of Zurich)</i>	
Import2vec: Learning Embeddings for Software Libraries	18
<i>Bart Theeten (Nokia Bell Labs, Belgium), Frederik Vandeputte (Nokia Bell Labs, Belgium), and Tom Van Cutsem (Nokia Bell Labs, Belgium)</i>	
Semantic Source Code Models Using Identifier Embeddings	29
<i>Vasiliki Efstathiou (Athens University of Economics and Business) and Diomidis Spinellis (Athens University of Economics and Business)</i>	

### Defect Prediction and Testing (Part 1)

DeepJIT: An End-to-End Deep Learning Framework for Just-in-Time Defect Prediction	34
<i>Thong Hoang (Singapore Management University), Hoa Khanh Dam (University of Wollongong), Yasutaka Kamei (Kyushu University), David Lo (Singapore Management University), and Naoyasu Ubayashi (Kyushu University)</i>	

Lessons Learned from Using a Deep Tree-Based Model for Software Defect Prediction in Practice .46.....	
<i>Hoa Khanh Dam (University of Wollongong), Trang Pham (Deakin University), Shien Wee Ng (University of Wollongong), Truyen Tran (Deakin University), John Grundy (Monash University), Aditya Ghose (University of Wollongong), Taeksu Kim (Samsung Electronics), and Chul-Joo Kim (Samsung Electronics)</i>	
Empirical Study in using Version Histories for Change Risk Classification .58.....	
<i>Max Kiehn (Advanced Micro Devices, Inc.), Xiangyi Pan (Advanced Micro Devices, Inc.), and Fatih Camci (Amazon)</i>	
Snoring: A Noise in Defect Prediction Datasets .63.....	
<i>Aalok Ahluwalia (California Polytechnic State University San Luis Obispo), Davide Falessi (California Polytechnic State University San Luis Obispo), and Massimiliano Di Penta (University of Sannio)</i>	

## Representations for Mining (Part 2)

Exploring Word Embedding Techniques to Improve Sentiment Analysis of Software Engineering Texts .68.....	
<i>Eeshita Biswas (University of Delaware), K. Vijay-Shanker (University of Delaware), and Lori Pollock (University of Delaware)</i>	
Cleaning StackOverflow for Machine Translation .79.....	
<i>Musfiqur Rahman (Concordia University), Peter Rigby (Concordia University), Dharani Palani (Concordia University), and Tien Nguyen (The University of Texas at Dallas)</i>	
Predicting Good Configurations for GitHub and Stack Overflow Topic Models .84.....	
<i>Christoph Treude (University of Adelaide) and Markus Wagner (University of Adelaide)</i>	

## Defect Prediction and Testing (Part 2)

A Dataset of Parametric Cryptographic Misuses .96.....	
<i>Anna-Katharina Wickert (Technische Universität Darmstadt), Michael Reif (Technische Universität Darmstadt), Michael Eichberg (Technische Universität Darmstadt), Anam Dodhy (Technische Universität Darmstadt), and Mira Mezini (Technische Universität Darmstadt)</i>	
Does UML Modeling Associate with Lower Defect Proneness?: A Preliminary Empirical Investigation .101.....	
<i>Adithya Raghuraman (Carnegie Mellon University), Truong Ho-Quang (Chalmers Gothenburg University), Michel R. V. Chaudron (Chalmers Gothenburg University), Alexander Serebrenik (Eindhoven University of Technology), and Bogdan Vasilescu (Carnegie Mellon University)</i>	
STRAIT: A Tool for Automated Software Reliability Growth Analysis .105.....	
<i>Stanislav Chren (Masaryk University), Radoslav Micko (Masaryk University), Barbora Buhnova (Masaryk University), and Bruno Rossi (Masaryk University)</i>	
A Data Set of Program Invariants and Error Paths .111.....	
<i>Dirk Beyer (LMU Munich)</i>	

Test Coverage in Python Programs .116.....	
	<i>Hongyu Zhai (University of California, Davis), Casey Casalnuovo (University of California, Davis), and Prem Devanbu (University of California, Davis)</i>
On the Effectiveness of Manual and Automatic Unit Test Generation: Ten Years Later .121.....	
	<i>Domenico Serra (University of Salerno), Giovanni Grano (University of Zurich), Fabio Palomba (University of Zurich), Filomena Ferrucci (University of Salerno), Harald C. Gall (University of Zurich), and Alberto Bacchelli (University of Zurich)</i>

## Education Track Tutorial I

### Large-Scale Mining

Time Present and Time Past: Analyzing the Evolution of JavaScript Code in the Wild .126.....	
	<i>Dimitris Mitropoulos (Athens University of Economics and Business), Panos Louridas (Athens University of Economics and Business), Vitalis Salis (Greek Research and Technology Network), and Diomidis Spinellis (Athens University of Economics and Business)</i>
The Software Heritage Graph Dataset: Public Software Development Under One Roof .138.....	
	<i>Antoine Pietri (Inria), Diomidis Spinellis (Athens University of Economics and Business), and Stefano Zacchiroli (University Paris Diderot and Inria)</i>
World of Code: An Infrastructure for Mining the Universe of Open Source VCS Data .143.....	
	<i>Yuxing Ma (University of Tennessee), Chris Bogart (Carnegie Mellon University), Sadika Amreen (University of Tennessee), Russell Zaretzki (University of Tennessee), and Audris Mockus (University of Tennessee)</i>
Crossflow: A Framework for Distributed Mining of Software Repositories .155.....	
	<i>Dimitris Kolovos (University of York), Patrick Neubauer (University of York), Konstantinos Bampis (University of York), Nicholas Matragkas (University of York), and Richard Paige (University of York)</i>

### Energy and Economics

Recommending Energy-Efficient Java Collections .160.....	
	<i>Wellington de Oliveira Júnior (Federal University of Pernambuco), Renato Oliveira dos Santos (Federal University of Pernambuco), Fernando José Castor de Lima Filho (Federal University of Pernambuco), Benito Fernandes de Araújo Neto (Federal University of Pernambuco), and Gustavo Henrique Lima Pinto (Federal University of Pará)</i>
GreenHub Farmer: Real-World Data for Android Energy Mining .171.....	
	<i>Hugo Matalonga (Universidade do Minho, Portugal), Bruno Cabral (Universidade de Coimbra, Portugal; CISUC, Portugal), Fernando Castor (Universidade Federal de Pernambuco, Brasil), Marco Couto (Universidade do Minho, Portugal), Rui Pereira (Universidade da Beira Interior, Portugal), Simao Melo de Sousa (Universidade da Beira Interior, Portugal; LISP, Portugal), and Joao Paulo Fernandes (Universidade de Coimbra, Portugal; CISUC, Portugal)</i>

GreenSource: A Large-Scale Collection of Android Code, Tests and Energy Metrics .176.....	<i>Rui Rua (University of Minho), Marco Couto (HASLab/INESCTEC), and João Saraiva (HASLab/INESCTEC)</i>
Striking Gold in Software Repositories? An Econometric Study of Cryptocurrencies on GitHub .181.....	<i>Asher Trockman (University of Evansville), Rijnard van Tonder (Carnegie Mellon University), and Bogdan Vasilescu (Carnegie Mellon University)</i>
A Panel Data Set of Cryptocurrency Development Activity on GitHub .186.....	<i>Rijnard van Tonder (Carnegie Mellon University), Asher Trockman (University of Evansville), and Claire Le Goues (Carnegie Mellon University)</i>

## Mining Challenge

SOTorrent: Studying the Origin, Evolution, and Usage of Stack Overflow Code Snippets .191.....	<i>Sebastian Baltes (University of Trier), Christoph Treude (The University of Adelaide), and Stephan Diehl (University of Trier)</i>
Mining Rule Violations in JavaScript Code Snippets .195.....	<i>Uriel Ferreira Campos (Federal University of Pará), Guilherme Smethurst (Federal University of Pará), João Pedro Moraes (Federal University of Pará), Rodrigo Bonifácio (University of Brasília), and Gustavo Pinto (Federal University of Pará)</i>
Snakes in Paradise?: Insecure Python-Related Coding Practices in Stack Overflow .200.....	<i>Akond Rahman (North Carolina State University), Effat Farhana (North Carolina State University), and Nasif Imtiaz (North Carolina State University)</i>
Man vs Machine – A Study into Language Identification of Stack Overflow Code Snippets .205.....	<i>Jens Dietrich (Victoria University of Wellington), Markus Luczak-Roesch (Victoria University of Wellington), and Elroy Dalefield (Victoria University of Wellington)</i>
Python Coding Style Compliance on Stack Overflow .210.....	<i>Nikolaos Bafatakis (University College London), Niels Boecker (University College London), Wenjie Boon (University College London), Martín Cabello Salazar (University College London), Jens Krinke (University College London), Gazi Oznacar (University College London), and Robert White (University College London)</i>
Towards Mining Answer Edits to Extract Evolution Patterns in Stack Overflow .215.....	<i>Themistoklis Diamantopoulos (Aristotle University of Thessaloniki), Maria Ioanna Sifaki (Aristotle University of Thessaloniki), and Andreas Symeonidis (Aristotle University of Thessaloniki)</i>
Analyzing Comment-Induced Updates on Stack Overflow .220.....	<i>Abhishek Soni (Independent Researcher) and Sarah Nadi (University of Alberta)</i>
What Edits are Done on the Highly Answered Questions in Stack Overflow? An Empirical Study .225.....	<i>Xianhao Jin (Virginia Polytechnic Institute) and Francisco Servant (Virginia Polytechnic Institute)</i>

Can Duplicate Questions on Stack Overflow Benefit the Software Development Community? .230.....  
*Durham Abric (McGill University), Oliver E. Clark (McGill University),  
Matthew Caminiti (McGill University), Keheliya Gallaba (McGill  
University), and Shane McIntosh (McGill University)*

How Often and What StackOverflow Posts Do Developers Reference in Their GitHub Projects? .235.....  
*Saraj Singh Manes (Carleton University) and Olga Baysal (Carleton  
University)*

Characterizing Duplicate Code Snippets between Stack Overflow and Tutorials .240.....  
*Manziba Akanda Nishi (Virginia Commonwealth University), Agnieszka  
Ciborowska (Virginia Commonwealth University), and Kostadin Damevski  
(Virginia Commonwealth University)*

Challenges with Responding to Static Analysis Tool Alerts .245.....  
*Nasif Imtiaz (North Carolina State University), Akond Rahman (North  
Carolina State University), Effat Farhana (North Carolina State  
university), and Laurie Williams (North Carolina State University)*

Impact of Stack Overflow Code Snippets on Software Cohesion: A Preliminary Study .250.....  
*Mashal Ahmad (Lero and University College Dublin) and Mel Ó Cinnéide  
(University College Dublin)*

We Need to Talk About Microservices: an Analysis from the Discussions on StackOverflow .255.....  
*Alan Bandeira (State University of Ceara), Carlos Alberto Medeiros  
(State University of Ceara), Matheus Paixao (State University of  
Ceara), and Paulo Henrique Maia (State University of Ceara)*

What do Developers Know About Machine Learning: A Study of ML Discussions on StackOverflow .260.....  
*Abdul Ali Bangash (University of Alberta), Hareem Sahar (University of  
Alberta), Shaiful Chowdhury (University of Alberta), Alexander William  
Wong (University of Alberta), Abram Hindle (University of Alberta),  
and Karim Ali (University of Alberta)*

## APIs & Dependencies (Part 1)

Investigating Next Steps in Static API-Misuse Detection .265.....  
*Amann Sven (CQSE GmbH), Hoan Anh Nguyen (Amazon.com, Inc), Sarah Nadi  
(University of Alberta), Tien N. Nguyen (University of Texas-Dallas),  
and Mira Mezini (Technische Universität Darmstadt)*

Identifying Experts in Software Libraries and Frameworks Among GitHub Users .276.....  
*João Eduardo Montandon (Federal University of Minas Gerais), Luciana  
Lourdes Silva (Federal Institute of Minas Gerais), and Marco Tulio  
Valente (Federal University of Minas Gerais)*

Data-Driven Solutions to Detect API Compatibility Issues in Android: An Empirical Study .288.....  
*Simone Scalabrino (University of Molise), Gabriele Bavota (Università  
della Svizzera Italiana), Mario Linares-Vásquez (Universidad de los  
Andes), Michele Lanza (Università della Svizzera Italiana), and Rocco  
Oliveto (University of Molise)*

## Automatic Summarization

- Generating Commit Messages from Diffs using Pointer-Generator Network .299.....  
*Qin Liu (Tongji University), Zihe Liu (Tongji University), Hongming Zhu (Tongji University), Hongfei Fan (Tongji University), Bowen Du (The University of Warwick), and Yu Qian (Tongji University)*
- Automatically Generating Documentation for Lambda Expressions in Java .310.....  
*Anwar Alqaimi (University of Adelaide), Patanamon Thongtanunam (University of Melbourne), and Christoph Treude (University of Adelaide)*
- Extracting API Tips from Developer Question and Answer Websites .321.....  
*Shaohua Wang (New Jersey Institute of Technology), NhatHai Phan (New Jersey Institute of Technology), Yan Wang (Central University of Finance and Economics, China), and Yong Zhao (New Jersey Institute of Technology)*

## APIs & Dependencies (Part 2)

- The Emergence of Software Diversity in Maven Central .333.....  
*César Soto-Valero (KTH Royal Institute of Technology), Amine Benelallam (Univ Rennes, Inria, CNRS, IRISA), Nicolas Harrand (KTH Royal Institute of Technology), Olivier Barais (Univ Rennes, Inria, CNRS, IRISA), and Benoit Baudry (KTH Royal Institute of Technology)*
- The Maven Dependency Graph: A Temporal Graph-Based Representation of Maven Central .344.....  
*Amine Benelallam (IRISA, Univ. Rennes, Inria, CNRS), Nicolas Harrand (KTH Royal Institute of Technology), César Soto-Valero (KTH Royal Institute of Technology), Benoit Baudry (KTH Royal Institute of Technology), and Olivier Barais (IRISA, Univ. Rennes, Inria, CNRS)*
- Dependency Versioning in the Wild .349.....  
*Jens Dietrich (Victoria University of Wellington), David Pearce (Victoria University of Wellington), Jacob Stringer (Massey University), Amjed Tahir (Massey University), and Kelly Blincoe (University of Auckland)*
- Splitting APIs: An Exploratory Study of Software Unbundling .360.....  
*Anderson Severo de Matos (Federal University of Ceará), João Bosco Ferreira Filho (Federal University of Ceará), and Lincoln Souza Rocha (Federal University of Ceará)*

## Security

- Automated Software Vulnerability Assessment with Concept Drift .371.....  
*Triet Huynh Minh Le (The University of Adelaide), Bushra Sabir (The University of Adelaide), and Muhammad Ali Babar (The University of Adelaide)*
- A Manually-Curated Dataset of Fixes to Vulnerabilities of Open-Source Software .383.....  
*Serena Elisa Ponta (SAP Security Research), Henrik Plate (SAP Security Research), Antonino Sabetta (SAP Security Research), Michele Bezzi (SAP Security Research), and Cédric Dangremont (SAP)*



Negative Results on Mining Crypto-API Usage Rules in Android Apps .388.....	
	<i>Jun Gao (University of Luxembourg), Pingfan Kong (University of Luxembourg), Li Li (Monash University), Tegawendé F. Bissyandé (University of Luxembourg), and Jacques Klein (University of Luxembourg)</i>
A Dataset of Non-Functional Bugs .399.....	
	<i>Aida Radu (University of Alberta) and Sarah Nadi (University of Alberta)</i>
RmvDroid: Towards A Reliable Android Malware Dataset with App Metadata .404.....	
	<i>Haoyu Wang (Beijing University of Posts and Telecommunications), Junjun Si (Changan Communication Technology Co., LTD.), Hao Li (OrangeApk, Inc.), and Yao Guo (Peking University)</i>

## Collaboration & Communication (Part 1)

An Empirical Study of Multiple Names and Email Addresses in OSS Version Control Repositories .409.....	
	<i>Jiixin Zhu (Institute of Software, Chinese Academy of Sciences) and Jun Wei (Institute of Software, Chinese Academy of Sciences)</i>
Characterizing the Roles of Contributors in Open-Source Scientific Software Projects .421.....	
	<i>Reed Milewicz (Sandia National Laboratories), Gustavo Pinto (Federal University of Para), and Paige Rodeghero (Clemson University)</i>
git2net - Mining Time-Stamped Co-Editing Networks from Large git Repositories .433.....	
	<i>Christoph Gote (ETH Zürich), Ingo Scholtes (University of Zürich), and Frank Schweitzer (ETH Zürich)</i>

## Software Quality (part 1)

The Rise of Android Code Smells: Who is to Blame? .445.....	
	<i>Sarra Habchi (Inria Lille - University of Lille), Naouel Moha (Université du Québec à Montréal), and Romain Rouvoy (University of Lille)</i>
Assessing Diffusion and Perception of Test Smells in Scala Projects .457.....	
	<i>Jonas De Bleser (Vrije Universiteit Brussel), Dario Di Nucci (Vrije Universiteit Brussel), and Coen De Roover (Vrije Universiteit Brussel)</i>
Style-Analyzer: Fixing Code Style Inconsistencies with Interpretable Unsupervised Algorithms .468.....	
	<i>Vadim Markovtsev (source{d}), Waren Long (source{d}), Hugo Mougard (source{d}), Konstantin Slavnov (source{d}), and Egor Bulychev (source{d})</i>

## Collaboration & Communication (Part 2)

Can Issues Reported at Stack Overflow Questions be Reproduced? An Exploratory Study .479.....	
	<i>Saikat Mondal (University of Saskatchewan), Mohammad Masudur Rahman (University of Saskatchewan), and Chanchal K. Roy (University of Saskatchewan)</i>

Exploratory Study of Slack Q&A Chats as a Mining Source for Software Engineering Tools .490.....	
	<i>Preetha Chatterjee (University of Delaware), Kostadin Damevski (Virginia Commonwealth University), Lori Pollock (University of Delaware), Vinay Augustine (ABB Corporate Research), and Nicholas A. Kraft (ABB Corporate Research)</i>
Impacts of Daylight Saving Time on Software Development .502.....	
	<i>Junichi Hayashi (Osaka University), Yoshiki Higo (Osaka University), Shinsuke Matsumoto (Osaka University), and Shinji Kusumoto (Osaka University)</i>

## Software Quality (part 2)

A Large-Scale Study About Quality and Reproducibility of Jupyter Notebooks .507.....	
	<i>João Felipe Pimentel (Universidade Federal Fluminense), Leonardo Murta (Universidade Federal Fluminense), Vanessa Braganholo (Universidade Federal Fluminense), and Juliana Freire (New York University)</i>
Cross-Language Clone Detection by Learning Over Abstract Syntax Trees .518.....	
	<i>Daniel Perez (Imperial College London) and Shigeru Chiba (The University of Tokyo)</i>
SeSaMe: A Data Set of Semantically Similar Java Methods .529.....	
	<i>Marius Kamp (Friedrich-Alexander University Erlangen-Nürnberg (FAU), Germany), Patrick Kreuzer (Friedrich-Alexander University Erlangen-Nürnberg (FAU), Germany), and Michael Philippsen (Friedrich-Alexander University Erlangen-Nürnberg (FAU), Germany)</i>

## Education Track Tutorial II

### Traceability

Predicting Co-Changes between Functionality Specifications and Source Code in Behavior Driven Development .534.....	
	<i>Aidan Z.H. Yang (Queen's University, Canada), Daniel Alencar da Costa (University of Otago), and Ying Zou (Queen's University, Canada)</i>
Tracing Back Log Data to its Log Statement: From Research to Practice .545.....	
	<i>Daan Schipper (Adyen N.V.), Mauricio Aniche (Delft University of Technology), and Arie van Deursen (Delft University of Technology)</i>
Beyond GumTree: A Hybrid Approach to Generate Edit Scripts .550.....	
	<i>Junnosuke Matsumoto (Osaka University), Yoshiki Higo (Osaka University), and Shinji Kusumoto (Osaka University)</i>
The Impact of Systematic Edits in History Slicing .555.....	
	<i>Ryosuke Funaki (Tokyo Institute of Technology), Shinpei Hayashi (Tokyo Institute of Technology), and Motoshi Saeki (Tokyo Institute of Technology)</i>
Scalable Software Merging Studies with MERGANSER .560.....	
	<i>Moein Owahdi-Kareshk (University of Alberta) and Sarah Nadi (University of Alberta)</i>

## Building on Data

Standing on Shoulders or Feet? The Usage of the MSR Data Papers .565.....	
<i>Zoe Kotti (Athens University of Economics and Business) and Diomidis Spinellis (Athens University of Economics and Business)</i>	
Boa Meets Python: A Boa Dataset of Data Science Software in Python Language .577.....	
<i>Sumon Biswas (Iowa State University), Md Johirul Islam (Iowa State University), Yijia Huang (Iowa State University), and Hriday Rajan (Iowa State University)</i>	
A Benchmark of Data Loss Bugs for Android Apps .582.....	
<i>Oliviero Riganelli (University of Milano - Bicocca), Marco Mobilio (University of Milano - Bicocca), Daniela Micucci (University of Milano - Bicocca), and Leonardo Mariani (University of Milano - Bicocca)</i>	
RapidRelease - A Dataset of Projects and Issues on Github with Rapid Releases .587.....	
<i>Saket Dattatray Joshi (Indian Institute of Technology Tirupati) and Sridhar Chimalakonda (Indian Institute of Technology Tirupati)</i>	
ConPan: A Tool to Analyze Packages in Software Containers .592.....	
<i>Ahmed Zerouali (University of Mons), Valerio Cosentino (Bitergia – Spain), Gregorio Robles (Universidad Rey Juan Carlos), Jesus M. Gonzalez-Barahona (Universidad Rey Juan Carlos), and Tom Mens (University of Mons)</i>	
An Empirical History of Permission Requests and Mistakes in Open Source Android Apps .597.....	
<i>Gian Luca Scoccia (Gran Sasso Science Institute), Anthony Peruma (Rochester Institute of Technology), Virginia Pujols (Rochester Institute of Technology), Ben Christians (Rochester Institute of Technology), and Daniel Krutz (Rochester Institute of Technology)</i>	
<b>Author Index 603</b> .....	