2021 IEEE International Conference on Software Analysis, **Evolution and Reengineering** (SANER 2021)

Virtual Conference 9 – 12 March 2021



IEEE Catalog Number: CFP21102-POD ISBN:

978-1-7281-9631-2

Copyright © 2021 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:
 CFP21102-POD

 ISBN (Print-On-Demand):
 978-1-7281-9631-2

 ISBN (Online):
 978-1-7281-9630-5

ISSN: 1534-5351

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



2021 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER) SANER 2021

Table of Contents

Message from the General Chair and Program Co-Chairs	xv
Organizing Committee	xvii
Program Committee	xix
Steering Committee	xxi
Keynotes	
Research Track	
Towards Understanding Tool-Chain Bugs in the LLVM Compiler Infrastructure	1
Analyzing and Detecting Emerging Quality-Related Concerns Across OSS Defect Report	
Summaries	12
An Empirical Study for Common Language Features Used in Python Projects	24
Graph Neural Network Based Collaborative Filtering for API Usage Recommendation	36
MulCode: A Multi-Task Learning Approach for Source Code Understanding Deze Wang (National University of Defense Technology, China), Yue Yu (National University of Defense Technology, China), Shanshan Li (National University of Defense Technology, China), Wei Dong (National University of Defense Technology, China), Ji Wang (National University of Defense Technology, China), and Liao Qing (Harbin Institute of Technology (Shenzhen), China)	48
A Neural Question Answering System for Basic Questions about Subroutines Aakash Bansal (University of Notre Dame, USA), Zachary Eberhart (University of Notre Dame, USA), Lingfei Wu (IBM Research, USA), and Collin McMillan (University of Notre Dame, USA)	60

A Multi-Metric Ranking Approach for Library Migration Recommendations Hao He (Peking University, China), Yulin Xu (Peking University, China), Yixiao Ma (Peking University, China), Yifei Xu (Peking University, China), Guangtai Liang (Huawei Technologies Co., Ltd., China), and Minghui Zhou (Peking University, China)	.72
HashMTI: Scalable Mutation-Based Taint Inference with Hash Records Xiangdong Kong (National University of Defense Technology, China), Yong Tang (National University of Defense Technology, China; Hunan Normal University, China), Pengfei Wang (National University of Defense Technology, China), Shuning Wei (Hunan Normal University, China), and Tai Yue (National University of Defense Technology, China)	84
Is the Ground Truth Really Accurate? Dataset Purification for Automated Program Repair Deheng Yang (National University of Defense Technology, China), Yan Lei (Chongqing University, China), Xiaoguang Mao (National University of Defense Technology, China), David Lo (Singapore Management University, Singapore), Huan Xie (Chongqing University, China), and Meng Yan (Chongqing University, China)	. 96
OTA: An Operation-Oriented Time Allocation Strategy for Greybox Fuzzing	108
Leveraging Stack Overflow to Detect Relevant Tutorial Fragments of APIs Di Wu (Wuhan University), Xiao-Yuan Jing (Wuhan University, Guangdong University of Petrochemical Technology), Hongyu Zhang (The University of Newcastle), Yuming Zhou (Nanjing University), and Baowen Xu (Nanjing University)	119
Binary Level Toolchain Provenance Identification with Graph Neural Networks	131
Assessing Time-Based and Range-Based Strategies for Commit Assignment to Releases	142
XSnare: Application-Specific Client-Side Cross-Site Scripting Protection Jose Carlos Pazos (University of British Columbia), Jean-Sébastien Légaré (University of British Columbia), and Ivan Beschastnikh (University of British Columbia)	154
Combining Query Reduction and Expansion for Text-Retrieval-Based Bug Localization	166
A Survey-Based Empirical Evaluation of Bad Smells in LabVIEW Systems Models	177

DeepCon: Contribution Coverage Testing for Deep Learning Systems Zhiyang Zhou (Institute of Software, Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China), Wensheng Dou (Institute of Software, Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China; Institute of Software Technology, China), Jie Liu (Institute of Software, Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China), Chenxin Zhang (Institute of Software, Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China), Jun Wei (Institute of Software, Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China), and Dan Ye (Institute of Software, Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China)	185
FrameFix: Automatically Repairing Statically-Detected Directive Violations in Framework Applications	201
VIVA: Binary Level Vulnerability Identification via Partial Signature Yang Xiao (Institute of Information Engineering, China), Zhengzi Xu (Nanyang Technological University, Singapore), Weiwei Zhang (University of Science and Technology of China, China), Chendong Yu (Institute of Information Engineering, China), Longquan Liu (Institute of Information Engineering, China), Wei Zou (Institute of Information Engineering, China), Zimu Yuan (Institute of Information Engineering, China), Yang Liu (Nanyang Technological University, Singapore), Aihua Piao (Institute of Information Engineering, China), and Wei Huo (Institute of Information Engineering, China)	213
Assessing Code Clone Harmfulness: Indicators, Factors, and Counter Measures Bin Hu (Fudan University, China), Yijian Wu (Fudan University, China), Xin Peng (Fudan University, China), Jun Sun (Singapore Management University, Singapore), Nanjie Zhan (Fudan University, China), and Jun Wu (Fudan University, China)	225
The Usability (or Not) of Refactoring Tools	237
Javadoc Violations and Their Evolution in Open-Source Software Marcel Steinbeck (University of Bremen, Germany) and Rainer Koschke (University of Bremen, Germany)	249
Quality Assurance for Automated Commit Message Generation	260

Understanding and Facilitating the Co-Evolution of Production and Test Code Sinan Wang (Southern University of Science and Technology, China), Ming Wen (Huazhong University of Science and Technology, China), Yepang Liu (Southern University of Science and Technology, China), Ying Wang (Northeastern University, China), and Rongxin Wu (Xiamen University, China)	272
Leveraging Fault Localisation to Enhance Defect Prediction	284
On the Impact of Flaky Tests in Automated Program Repair	295
GhostBuster: Understanding and Overcoming the Pitfalls of Transient Execution Vulnerability Checkers	307
Automatic Detection of Five API Documentation Smells: Practitioners' Perspectives	318
Action Word Prediction for Neural Source Code Summarization	330
Two-Stage Attention-Based Model for Code Search with Textual and Structural Features	342
SQLIFIX: Learning Based Approach to Fix SQL Injection Vulnerabilities in Source Code	354

Call Frequency-Based Fault Localization	365
A Multi-Criteria Strategy for Redesigning Legacy Features as Microservices: An Industrial Case Study	377
Underproduction: An Approach for Measuring Risk in Open Source Software	388
A Large-Scale Empirical Study of Android App Decompilation	400
EnHMM: On the Use of Ensemble HMMs and Stack Traces to Predict the Reassignment of Bug Report Fields	411
On the Co-Evolution of ML Pipelines and Source Code - Empirical Study of DVC Projects	422
Summarizing Relevant Parts from Technical Videos	434
Empirical Analysis of Security Vulnerabilities in Python Packages	446
Plot2API: Recommending Graphic API from Plot via Semantic Parsing Guided Neural Network *Zeyu Wang (Chongqing University, China), Sheng Huang (Chongqing University, China), Zhongxin Liu (Zhejiang University, China), Meng Yan (Chongqing University, China), Xin Xia (Monash University, Australia), Bei Wang (Chongqing University, China), and Dan Yang (Chongqing University, China)	458

Understanding Code Reuse in Smart Contracts	.470
Yat-sen University, China), Yixin Zhang (Sun Yat-sen University, China), Yuan Huang (Sun Yat-sen University, China), and Zibin Zheng (Sun Yat-sen University, China)	
Android Compatibility Issue Detection Using API Differences	480
ERA Track	
Evaluating Bug Prediction Under Realistic Settings	491
Multi-Threshold Token-Based Code Clone Detection	496
Onboarding to Open Source Projects with Good First Issues: A Preliminary Analysis	501
Human-Centric Issues in eHealth App Development and Usage: A Preliminary Assessment	506
Enhancing Just-in-Time Defect Prediction Using Change Request-Based Metrics	511
Linking Developer Experience to Coding Style in Open-Source Repositories	. 516
Adaptive Immunity for Software: Towards Autonomous Self-Healing Systems	.521
How Do Users Answer MATLAB Questions on Q&A Sites? A Case Study on Stack Overflow and MathWorks Mahshid Naghashzadeh (Shiraz University, Iran), Amir Haghshenas (Shiraz University, Iran), Ashkan Sami (Shiraz University, Iran), and David Lo (Singapore Management University, Singapore)	526

Anti-Patterns in Modern Code Review: Symptoms and Prevalence	. 531
MSR4ML: Reconstructing Artifact Traceability in Machine Learning Repositories	536
Automatically Customizing Static Analysis Tools to Coding Rules Really Followed by Developers Yuki Ueda (Nara Institute of Science and Technology, Japan), Takashi Ishio (Nara Institute of Science and Technology, Japan), and Kenichi Matsumoto (Nara Institute of Science and Technology, Japan)	. 541
Using Structural and Semantic Information to Identify Software Components	546
Tool Track	
G-Repo: a Tool to Support MSR Studies on GitHub	. 551
Gasser	. 556
Carmen Coviello (University of Basilicata, Italy), Simone Romano (University of Bari, Italy), Giuseppe Scanniello (University of Basilicata, Italy), and Giuliano Antoniol (Polytechnique Montreal, Canada)	
Recording, Visualising and Understanding Developer Programming Behaviour	561
ClarchMap: API-Oriented Reconstruction of Architecture for Cloud-Based Systems	567
TinySpline: A Small, yet Powerful Library for Interpolating, Transforming, and Querying NURBS, B-Splines, and Bézier Curves	. 572

Makar: A Framework for Multi-Source Studies Based on Unstructured Data	7
Industry Track	
Practitioners' Testimonials about Software Testing	32
Technical Debt in Large-Scale Distributed Projects: An Industrial Case Study	0'
Mining for Graph-Based Library Usage Patterns in COBOL Systems	15
A Testing Approach While Re-Engineering Legacy Systems: An Industrial Case Study)C
How do Technical Debt Payment Practices Relate to the Effects of the Presence of Debt	
Items in Software Projects?	
RENE Track	
How Explicit Feature Traces Did Not Impact Developers' Memory Jacob Krüger (Otto-von-Guericke University Magdeburg, Germany), Gül Çalıklı (University of Zurich, Switzerland), Thorsten Berger (Ruhr-University Bochum, Germany & Chalmers, University of Gothenburg, Sweden), and Thomas Leich (Harz University of Applied Sciences Wernigerode & METOP GmbH Magdeburg, Germany)	.С

Will Data Influence the Experiment Results?: A Replication Study of Automatic Identification of Decisions
Liming Fu (Wuhan University, China), Peng Liang (Wuhan University, China), Xueying Li (Wuhan University, China), and Chen Yang (IBO Technology (Shenzhen) Co., Ltd., China)
Workshops
Regression Test Selection Tool for Python in Continuous Integration Process 618 Eero Kauhanen (University of Helsinki, Finland), Jukka K. Nurminen (University of Helsinki, Finland), Tommi Mikkonen (University of Helsinki, Finland), and Matvey Pashkovskiy (F-Secure Corporation, Finland)
An Analysis of Testing Scenarios for Automated Driving Systems
A Case Study on the Evolution of Configuration Options of a Highly-Configurable Software System
AI-Based Enhancement of Test Models in an Industrial Model-Based Testing Tool
Research Progress of Flaky Tests
Representation vs. Model: What Matters Most for Source Code Vulnerability Detection
Bug Question Answering with Pretrained Encoders
Characterizing Duplicate Bugs: An Empirical Analysis
Blockchain in Agri-Food Traceability Systems: a Model Proposal for a Typical Italian Food Product
Drone-Based Risk Management of Autonomous Systems Using Contracts and Blockchain

Lightweight Extension of an Execution Environment for Safer Function Calls in Solidity/Ethereum Virtual Machine Smart Contracts Sooyeon Lee (Chungnam National University, Republic of Korea) and Eun-Sun Cho (Chungnam National University, Republic of Korea)	689
Raising Sustainability Awareness in Agile Blockchain-Oriented Software Engineering	. 696
Analysis of Source Code Duplication in Ethreum Smart Contracts	. 701
Smart-Graph: Graphical Representations for Smart Contract on the Ethereum Blockchain	. 708
Investigation of Blockchain Cryptocurrencies' Price Movements Through Deep Learning: A Comparative Analysis	715
Blockchain and Contact Tracing Applications for COVID-19: The Opportunity and The Challenges	. 72 3
Author Index	731