2023 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER 2023)

Macao, China 21-24 March 2023

Pages 1-484



IEEE Catalog Number: CFP23102-POD **ISBN:**

978-1-6654-5279-3

Copyright © 2023 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:	CFP23102-POD
ISBN (Print-On-Demand):	978-1-6654-5279-3
ISBN (Online):	978-1-6654-5278-6
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



2023 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER) **SANER 2023**

Table of Contents

Message from the General Chair and Program Co-Chairs	xix
Organizing Committee	xxi
Program Committee	xxiii
Steering Committee	
Keynotes	xxix
Invited Talks	

Research Track

An Empirical Study of Smart Contract Decompilers
Understanding the Archived Projects on GitHub
Towards Understanding the Impacts of Textual Dissimilarity on Duplicate Bug Report 25 Detection 25 Sigma Jahan (Dalhousie University) and Mohammad Masudur Rahman 25 (Dalhousie University) 25
A Hierarchical Topical Modeling Approach for Recommending Repair of Quality Bugs
Solder: Retrofitting Legacy Code with Cross-Language Patches
Pseudocode to Code Based on Adaptive Global and Local Information

Identifying Emergent Leadership in Open Source Software Projects Based on Communication Styles	3
Yuekai Huang (Institute of Software Chinese Academy of Sciencess, China; University of Chinese Academy of Sciences, China), Ye Yang (Stevens Institute of Technology, USA), Junjie Wang (Institute of Software Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China), Wei Zheng (Stevens Institute of Technology, USA), and Qing Wang (Institute of Software, Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China; University of Sciences, China)	
NeuralCCD Integrating Multiple Features for Neural Coincidental Correctness Detection	5
 Studying and Complementing the Use of Identifiers in Logs	7
Automatic Identification of Crash-Inducing Smart Contracts	3
 MulCS: Towards a Unified Deep Representation for Multilingual Code Search)
Efficient Pattern-Based Static Analysis Approach via Regular-Expression Rules	<u>)</u>
Intelligent Constraint Classification for Symbolic Execution	ł
 Cross-Language Call Graph Construction Supporting Different Host Languages	5

Topic Recommendation for GitHub Repositories: How Far Can Extreme Multi-Label Learning Go? 167
Ratnadira Widyasari (Singapore Management University), Zhipeng Zhao (Singapore Management University), Thanh Le Cong (Singapore Management University), Hong Jin Kang (Singapore Management University), and David Lo (Singapore Management University)
An Empirical Comparison of Model-Agnostic Techniques for Defect Prediction Models 179 Gichan Lee (Hanyang University, Republic of Korea) and Scott Uk-Jin Lee (Hanyang University, Republic of Korea)
Security Code Recommendations for Smart Contract
On the Effect of Transitivity and Granularity on Vulnerability Propagation in the Maven Ecosystem 201 Amir M. Mir (Delft University of Technology, The Netherlands), Mehdi 201 Keshani (Delft University of Technology, The Netherlands), and Sebastian Proksch (Delft University of Technology, The Netherlands)
CLAWSAT: Towards Both Robust and Accurate Code Models
Efficient Generation of Floating-Point Inputs for Compiler-Induced Variability
In-Situ Concolic Testing of JavaScript
Learning Program Representations with a Tree-Structured Transformer
Extending Source Code Pre-Trained Language Models to Summarise Decompiled Binaries

Multi-Modal API Recommendation
Ivana Clairine Irsan (Singapore Management University, Singapore), Ting Zhang (Singapore Management University, Sinapore), Ferdian Thung (Singapore Management University, Singapore), Kisub Kim (Singapore Management University, Singapore), and David Lo (Singapore Management University, Singapore)
 Analysis and Propagation of Feature Revisions in Preprocessor-Based Software Product Lines 284 Gabriela K. Michelon (Johannes Kepler University Linz, Austria), Wesley K. G. Assunção (Johannes Kepler University Linz, Austria), Paul Grünbacher (Johannes Kepler University Linz, Austria), and Alexander Egyed (Johannes Kepler University Linz, Austria)
D-ACT: Towards Diff-Aware Code Transformation for Code Review Under a Time-Wise Evaluation
296 Chanathip Pornprasit (Monash University, Australia), Chakkrit Tantithamthavorn (Monash University, Australia), Patanamon Thongtanunam (The University of Melbourne, Australia), and Chunyang Chen (Monash University, Australia)
UDA-DP: Unsupervised Domain Adaptation for Software Defect Prediction
Slice-Based Code Change Representation Learning
Improving Fine-Tuning Pre-Trained Models on Small Source Code Datasets via Variational Information Bottleneck 331 Jiaxing Liu (Fudan University), Chaofeng Sha (Fudan University), and 331 Xin Peng (Fudan University) 331
Documentation-Guided API Sequence Search Without Worrying About the Text-API Semantic Gap 343
Hongwei Wei (Harbin Institute of Technology, China), Xiaohong Su (Harbin Institute of Technology, China), Weining Zheng (Harbin Institute of Technology, China), and Wenxin Tao (Harbin Institute of Technology, China)
A Comprehensive Study of WebAssembly Runtime Bugs
 Automated Image Reduction for Explaining Black-box Classifiers

 MixCode: Enhancing Code Classification by Mixup-Based Data Augmentation
Enhancing Comprehension and Navigation in Jupyter Notebooks with Static Analysis
CSSAM: Code Search via Attention Matching of Code Semantics and Structures
Towards a Large-Scale Empirical Study of Python Static Type Annotations414Xinrong Lin (University of Science and Technology of China, China),8aojian Hua (University of Science and Technology of China, China),Yang Wang (University of Science and Technology of China, China),Yang Wang (University of Science and Technology of China, China),Yang Wang (University of Science and Technology of China, China),Yang Wang (University of Science and Technology of China, China),
Understanding the Impact of Quantum Noise on Quantum Programs
Optimizing Continuous Integration by Dynamic Test Proportion Selection
A Multi-Step Learning Approach to Assist Code Review
 Evaluating Learning-to-Rank Models for Prioritizing Code Review Requests using Process Simulation
Characterize Software Release Notes of GitHub Projects: Structure, Writing Style, and Content

Towards Privacy Preserving Cross Project Defect Prediction with Federated Learning	
 Automatic Code Annotation Generation Based on Heterogeneous Graph Structure	
Brigadier: A Datalog-Based IAST Framework for Node.js Applications	
 How Dynamic Features Affect API Usages? An Empirical Study of API Misuses in Python Programs	•
Test Input Selection for Deep Neural Network Enhancement Based on Multiple-Objective Optimization	•
IRP2API: Automated Mapping of Cyber Security Incident Response Plan to Security Tools' APIs	
Detecting JavaScript Transpiler Bugs with Grammar-Guided Mutation	,
Towards Understanding Fixes of SonarQube Static Analysis Violations: A Large-Scale Empirical Study	,

Self-Adaptive Ensemble-Based Approach for Software Effort Estimation Suyash Shukla (Indian Institute of Technology Roorkee, India) and Sandeep Kumar (Indian Institute of Technology Roorkee, India)	581
Boosting Compiler Testing via Eliminating Test Programs with Long-Execution-Time Jiangchang Wu (Nanjing University, China), Yibiao Yang (Nanjing University, China), and Yuming Zhou (Nanjing University, China)	593
Prediction of Vulnerability Characteristics Based on Vulnerability Description and Prompt Learning Xiangwei Li (University of Science and Technology of China, China), Xiaoning Ren (University of Science and Technology of China, China), Yinxing Xue (University of Science and Technology of China, China), Zhenchang Xing (CSIRO's Data61, Australia; Australian National University, Australia), and Jiamou Sun (Australian National University, Australia)	604
 Architecture Decisions in AI-Based Systems Development: An Empirical Study Beiqi Zhang (Wuhan University, China; Hubei Luojia Laboratory, China), Tianyang Liu (University of California San Diego, USA), Peng Liang (Wuhan University, China; Hubei Luojia Laboratory, China), Chong Wang (Wuhan University, China; Hubei Luojia Laboratory, China), Mojtaba Shahin (RMIT University, Australia), and Jiaxin Yu (Wuhan University, China; Hubei Luojia Laboratory, China) 	616
Self-Admitted Library Migrations in Java, JavaScript, and Python Packaging Ecosystems: A Comparative Study Haiqiao Gu (Peking University, China), Hao He (Peking University, China), and Minghui Zhou (Peking University, China)	627
RGDroid: Detecting Android Malware with Graph Convolutional Networks Against Structural Attack	639
 Exploring and Repairing Gender Fairness Violations in Word Embedding-Based Sentiment Analysis Model Through Adversarial Patches Lin Sze Khoo (Monash University Malaysia), Jia Qi Bay (Monash University Malaysia), Ming Lee Kimberly Yap (Monash University Malaysia), Mei Kuan Lim (Monash University Malaysia), Chun Yong Chong (Monash University Malaysia), Zhou Yang (Singapore Management University), and David Lo (Singapore Management University) 	651

ERA Track

An Adaptive Partition-Based Approach for Adaptive Random Testing on Real Programs	58
WasmSlim: Optimizing WebAssembly Binary Distribution via Automatic Module Splitting	73
Large Language Models: The Next Frontier for Variable Discovery Within Metamorphic 67 Testing? 67 Christos Tsigkanos (University of Bern, Switzerland), Pooja Rani 67	78
(University of Zurich, Switzerland), Sebastian Müller (Humboldt-Universität zu Berlin, Germany), and Timo Kehrer (University of Bern, Switzerland)	
Combining Contexts from Multiple Sources for Documentation-Specific Code Example Generation	22
Junaed Younus Khan (University of Calgary, Canada) and Gias Uddin (University of Calgary, Canada)	55
CGORewritter: A Better way to use C Library in Go	38
An Empirical Investigation on the Performance of Domain Adaptation for T5 Code Completion69 Daisuke Fukumoto (Nara Institute of Science and Technology, Japan), Yutaro Kashiwa (Nara Institute of Science and Technology, Japan), Toshiki Hirao (Nara Institute of Science and Technology, Japan), Kenji Fujiwara (Tokyo City University, Japan), and Hajimu Iida (Nara Institute of Science and Technology, Japan)	¥3
Weak Labelling for File-Level Source Code Classification) 8
Towards Automated Code Reviews: Does Learning Code Structure Help? 70 Hong Yi Lin (The University of Melbourne, Australia) and Patanamon 70 Thongtanunam (The University of Melbourne, Australia))3
How Robust Is a Large Pre-Trained Language Model for Code Generation? A Case on Attacking GPT2	าย
Rui Zhu (Yunnan University, China) and Cunming Zhang (Yunnan University, China)	,0
JTestMigBench and JTestMigTax: A Benchmark and Taxonomy for unit test Migration71 Ajay Kumar Jha (North Dakota State University, USA), Mohayeminul Islam (University of Alberta, Canada), and Sarah Nadi (University of Alberta, Canada)	13

TADL: Fault Localization with Transformer-Based Anomaly Detection for Dynamic Microservice

)	 J	
Systems	5	 	 	 	 	718
-						

Yuewei Li (Beijing University of Posts and Telecommunications, China), Yan Lu (Beijing University of Posts and Telecommunications, China), Jingyu Wang (Beijing University of Posts and Telecommunications, China), Qi Qi (Beijing University of Posts and Telecommunications, China), Jing Wang (Beijing University of Posts and Telecommunications, China), Yingying Wang (China Mobile Research Institute, China), and

Jianxin Liao (Beijing University of Posts and Telecommunications, China)

Tool Demo Track

TemLock: A Lightweight Template-Based Approach for Fixing Deadlocks Caused by ReentrantLock Lili Bo (Yangzhou University, China), Yanchi Yuan (Yangzhou	.723
University, China), Xiaobing Sun (Yangzhou University, China), Hao Xie (Yangzhou University, China), and Bin Li (Yangzhou University, China)	
HFCommunity: A Tool to Analyze the Hugging Face Hub Community Adem Ait (IN3 - UOC, Spain), Javier Luis Cánovas Izquierdo (IN3 - UOC, Spain), and Jordi Cabot (IN3 – UOC, ICREA, Spain)	728
Multi-Layer Observability for Fault Localization in Microservices Based Systems Rupashree Rangaiyengar (Indian Institute of Science, India), Raghavan Komondoor (Indian Institute of Science, India), and Raveendra Kumar Medicherla (TCS Research, India)	733
xVMP: An LLVM-Based Code Virtualization Obfuscator Xuangan Xiao (Shanghai Jiao Tong University, China), Yizhuo Wang (Shanghai Jiao Tong University, China), Yikun Hu (Shanghai Jiao Tong University, China), and Dawu Gu (Shanghai Jiao Tong University, China)	738
Naturalistic Static Program Analysis Mohammad Mehdi Pourhashem Kallehbasti (University of Science and Technology of Mazandaran, Iran) and Mohammad Ghafari (TU Clausthal, Germany)	743
μDetector: Automated Intrusion Detection for Microservices José Flora (University of Coimbra, Portugal), Miguel Teixeira (University of Coimbra, Portugal), and Nuno Antunes (University of Coimbra, Portugal)	748
 WasmA: A Static WebAssembly Analysis Framework for Everyone Florian Breitfelder (Technical University of Darmstadt, Germany), Tobias Roth (Technical University of Darmstadt, Germany), Lars Baumgärtner (Technical University of Darmstadt, Germany), and Mira Mezini (Technical University of Darmstadt, Germany) 	753

ISTA: Automatic Test Case Generation and Optimization for Intelligent Systems Based on Coverage Analysis
Industry Track
 Environment-Aware Testing for DNN-Based Smart-Home WiFi Sensing Systems
Leveraging and Evaluating Automatic Code Summarization for JPA Program Comprehension 768 Richard Mayer (Software Competence Center Hagenberg, Austria), Michael Moser (Software Competence Center Hagenberg, Austria), and Verena Geist (Software Competence Center Hagenberg, Austria)

An Experience Report on Regression-Free Repair of Deep Neural Network Model	778
Takao Nakagawa (Fujitsu Limited, Japan), Susumu Tokumoto (Fujitsu	
Limited, Japan), Shogo Tokui (Fujitsu Limited, Japan), and Fuyuki	
Ishikawa (National Institute of Informatics, Japan)	

RENE Track

Negative Results of Fusing Code and Documentation for Learning to Accurately Identify Sensitive Source and Sink Methods: An Application to the Android Framework for Data Leak	
	783
Jordan Samhi (University of Luxembourg, Luxembourg), Maria Kober	
(Fraunhofer Institute for Secure Information Technology, Germany),	
Abdoul Kader Kabore (University of Luxembourg, Luxembourg), Števen	
Arzt (Fraunhofer Institute for Secure Information Technology,	
Germany), Tegawendé F. Bissyandé (University of Luxembourg,	
Luxembourg), and Jacques Klein (University of Luxembourg, Luxembourg)	

 MLinter: Learning Coding Practices from Examples – Dream or Reality?	795
Revisiting Inter-Class Maintainability Indicators	305
Large-Scale Evaluation of Method-Level Bug Localization with FinerBench4BL	315
Technical Debt Diffuseness in the Apache Ecosystem: A Differentiated Replication	325
Evaluating and Improving Transformers pre-Trained on ASTs for Code Completion	334

Mezini	(TU	Darmstadt)

Journal First Track

Extended Abstract of A Comparative Study and Analysis of Developer Communications on Slack and Gitter Esteban Parra (Belmont University, USA), Mohammad Alahmad (University of Jeddah, Saudi Arabia), Ashley Ellis (Florida State University, USA), and Sonia Haiduc (Department of Computer Science, USA)	. 845
Extended Abstract of E-SC4R: Explaining Software Clustering for Remodularisation Alvin Jian Jia Tan (Monash University, Malaysia), Chun Yong Chong (Monash University, Malaysia), and Aldeida Aleti (Monash University, Australia)	. 846
Extended Abstract of SeCNN: A Semantic CNN Parser for Code Comment Generation Zheng Li (Beijing University of Chemical Technology, China), Yonghao Wu (Beijing University of Chemical Technology, China), Bin Peng (Beijing University of Chemical Technology, China), Xiang Chen (Nantong University, China; Nanjing University of Aeronautics and Astronautics, China; Beijing University of Chemical Technology, China), Zeyu Sun (Peking University, China), Yong Liu (Beijing University of Chemical Technology, China), and Deli Yu (Beijing University of Chemical Technology, China)	848

 Extended Abstract of Graph4Web: A Relation-Aware Graph Attention Network for Web Service Classification)
 Extended Abstract of Combine Sliced Joint Graph with Graph Neural Networks for Smart Contract Vulnerability Detection	1
 Extended Abstract of Candidate Test Set Reduction for Adaptive Random Testing: An Overheads Reduction Technique	3
-	_
Blockchain Oracles for Document Certification: A Case Study	5
Analysis of Users' Most Discussed Topics and Trends on Blockchain Technologies and Smart Contracts	5
An Optimized Concurrent Proof of Authority Consensus Protocol	1
An Analysis of the Oracles used in Ethereum's Blockchain	3
Investigating Software Testing and Maintenance of Open-Source Distributed Ledger	5
Detection of Test Smells with Basic Language Analysis Methods and its Evaluation	7

Bug or not Bug? Analysing the Reasons Behind Metamorphic Relation Violations
A Comparative Study of Code Clone Genealogies in Test Code and Production Code
Insights from Building a GUI Testing Tool
Deep Reinforcement Learning Guided Decision Tree Learning For Program Synthesis
EcoDialTest: Adaptive Mutation Schedule for Automated Dialogue Systems Testing
Interpretable Software Defect Prediction Incorporating Multiple Rules
A Cross-Language Name Binding Recognition and Discrimination Approach for Identifiers
 TAG: UML Activity Diagram Deeply Supervised Generation from Business Textural Specification
Multimodal Sentiment Analysis Under Modality Deficiency with Prototype-Augmentation in 962 Software Engineering 962 Baolei Wang (Yunnan University, China), Xuan Zhang (Yunnan University, China), Kunpeng Du (Yunnan University, China), Chen Gao (Yunnan University, China), and Linyu Li (Yunnan University, China)
SICUP: A Comment Updating Approach Based on Structural Information

ASTHGCN: Adaptive Spatio-Temporal Hypergraph Convolutional Network for Traffic Forecasting.... 972

Chao Zhu (Yunnan University, China), Jing Chen (Yunnan Highway Network Toll Management Co., Ltd, China), Rui Zhu (Yunnan University, China), Zhengqiong Wang (Yunnan Highway Network Toll Management Co., Ltd, China), Shihan Liu (Yunnan Highway Network Toll Management Co., Ltd, China), and Jishu Wang (Yunnan University, China)

Author Index