2023 IEEE 23rd International Working Conference on Source Code Analysis and Manipulation (SCAM 2023)

Bogota, Colombia
2 – 3 October 2023

IEEE Catalog Number: CFP23SRC-POD
ISBN: 979-8-3503-0507-4

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:	CFP23SRC-POD
ISBN (Print-On-Demand):	979-8-3503-0507-4
ISBN (Online):	979-8-3503-0506-7
ISSN:	1942-5430

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 23rd International Working Conference on Source Code Analysis and Manipulation (SCAM)

SCAM 2023

Table of Contents

Message from the General Chair and Program Co-Chairs ..... **ix**
Organizing Committee ..... **xi**
Program Committee ..... **xii**
Steering Committee ..... **xv**
Reviewers ..... **xvi**
Keynotes ..... **xvii**
Sponsors and Supporters ..... **xix**

Research Track

Applying the Universal Version History Concept to Help De-Risk Copy-Based Code Reuse 1
David Reid (University of Tennessee Knoxville, USA) and Audris Mockus (University of Tennessee Knoxville, USA)

Automating Test-Specific Refactoring Mining: A Mixed-Method Investigation 13
Luana Martins (Federal University of Bahia, Brazil), Heitor Costa (Federal University of Lavras, Brazil), Márcio Ribeiro (Federal University of Alagoas, Brazil), Fabio Palomba (University of Salerno, Italy), and Ivan Machado (Federal University of Bahia, Brazil)

Behind Developer Contributions on Conflicting Merge Scenarios 25
Gustavo Vale (Saarland University, Germany), Eduardo Fernandes (University of Southern Denmark, Denmark), Eduardo Figueiredo (Federal University of Minas Gerais, Brazil), and Sven Apel (Saarland University, Germany)

Calibrating Deep Learning-Based Code Smell Detection Using Human Feedback 37
Himesh Nandani (Dalhousie University, Canada), Mootez Saad (Dalhousie University, Canada), and Tushar Sharma (Dalhousie University, Canada)

Change Pattern Detection for Optimising Incremental Static Analysis 49
Cindy Wauters (Vrije Universiteit Brussel, Belgium), Jens Van der Plas (Vrije Universiteit Brussel, Belgium), Quentin Stiévenart (Université du Québec à Montréal, Canada), and Coen De Roover (Vrije Universiteit Brussel, Belgium)

CIGAR: Contrastive Learning for GitHub Action Recommendation 61
Jiangnan Huang (Radboud University, The Netherlands) and Bin Lin (Radboud University, The Netherlands)

Do Code Quality and Style Issues Differ Across (Non-)Machine Learning Notebooks? Yes! 72
Md Saeed Siddik (University of Alberta, Canada) and Cor-Paul Bezemer (University of Alberta, Canada)

Evolutionary Feature Dependencies: Analyzing Feature Co-Changes in C Systems 84
Sandro Schulze (Anhalt University of Applied Sciences, Germany), Phillipp Engelke (Otto-von-Guericke University Magdeburg, Germany), and Jacob Krüger (Eindhoven University of Technology, The Netherlands)

Explaining Transformer-Based Code Models: What Do They Learn? When They Do Not Work? 96
Ahmad Haji Mohammadkhani (University of Calgary, Canada), Chakkrit Tantithamthavorn (Monash University, Australia), and Hadi Hemmati (York University, Canada)

Generating Understandable Unit Tests Through End-to-End Test Scenario Carving 107
Amirhossein Deljouyi (Delft University of Technology, The Netherlands) and Andy Zaidman (Delft University of Technology, The Netherlands)

How They Relate and Leave: Understanding Atoms of Confusion in Open-Source Java Projects 119
Oton Pinheiro (Federal University of Ceará, Brazil), Lincoln Rocha (Federal University of Ceará, Brazil), and Windson Viana (Federal University of Ceará, Brazil)

How Well Can Masked Language Models Spot Identifiers That Violate Naming Guidelines? 131
Johannes Villmow (RheinMain University of Applied Sciences, Germany), Viola Campos (RheinMain University of Applied Sciences, Germany), Jean Petry (RheinMain University of Applied Sciences, Germany), Amine Abbad-Andaloussi (University of St. Gallen, Switzerland), Adrian Ulges (RheinMain University of Applied Sciences, Germany), and Barbara Weber (University of St. Gallen, Switzerland)

Leveraging User-Defined Identifiers for Counterfactual Data Generation in Source Code Vulnerability Detection 143
Hongyu Kuang (National Key Laboratory of Science and Technology on Information System Security, Systems Engineering Institute, Academy of Military Sciences, China), Feng Yang (National Key Laboratory of Science and Technology on Information System Security, Systems Engineering Institute, Academy of Military Sciences, China), Long Zhang (National Key Laboratory of Science and Technology on Information System Security, Systems Engineering Institute, Academy of Military Sciences, China), Gaigai Tang (National Key Laboratory of Science and Technology on Information System Security, Systems Engineering Institute, Academy of Military Sciences, China), and Lin Yang (National Key Laboratory of Science and Technology on Information System Security, Systems Engineering Institute, Academy of Military Sciences, China)

PASD: A Performance Analysis Approach Through the Statistical Debugging of Kernel Events 151
Mohammed Adib Khan (Brock University, Canada), Morteza Noferesti (Brock University, Canada), and Naser Ezzati-Jivan (Brock University, Canada)

PTLVD:Program Slicing and Transformer-Based Line-Level Vulnerability Detection System 162
Tao Peng (Wuhan Textile University, China), Shixu Chen (Wuhan Textile University, China), Fei Zhu (Wuhan Textile University, China), Junwei Tang (Wuhan Textile University, China), Junping Liu (Wuhan Textile University, China), and Xinrong Hu (Wuhan Textile University, China)

Quality Assurance Awareness in Open Source Software Projects on GitHub 174
Ali Khatami (Delft University of Technology, The Netherlands) and Andy Zaidman (Delft University of Technology, The Netherlands)

Symbolic Execution to Detect Semantic Merge Conflicts 186
Ward Muylaert (Vrije Universiteit Brussel, Belgium), Johannes Härtel (Vrije Universiteit Brussel, Belgium), and Coen De Roover (Vrije Universiteit Brussel, Belgium)

The Docker Hub Image Inheritance Network: Construction and Empirical Insights 198
Ruben Opdebeeck (Vrije Universiteit Brussel, Belgium), Jonas Lesy (Vrije Universiteit Brussel, Belgium), Ahmed Zerouali (Vrije Universiteit Brussel, Belgium), and Coen De Roover (Vrije Universiteit Brussel, Belgium)

Unboxing Default Argument Breaking Changes in Scikit Learn 209
João Eduardo Montandon (Universidade Federal de Minas Gerais, Brazil), Luciana Lourdes Silva (Instituto Federal de Minas Gerais, Brazil), Cristiano Politowski (École de Technologie Supérieure, Canada), Ghizlane El Boussaidi (École de Technologie Supérieure, Canada), and Marco Tulio Valente (Universidade Federal de Minas Gerais, Brazil)

Using the TypeScript Compiler to fix Erroneous Node.js Snippets 220
Brittany Reid (The University of Adelaide, Australia), Christoph Treude (The Unversity of Melbourne, Australia), and Markus Wagner (Monash University, Australia)

When to Let the Developer Guide: Trade-offs Between Open and Guided Test Amplification 231
Carolin Brandt (Delft University of Technology), Danyao Wang (Delft University of Technology), and Andy Zaidman (Delft University of Technology)

Engineering Track

Enabling Go Program Analysis in Rascal 242
Luke Swearngan (Appalachian State University, USA) and Mark Hills (Appalachian State University, USA)

On Developing and Improving Tools for Architecture-Smell Tracking in Java Systems 248
Philipp Gnoyke (Otto-von-Guericke University, Germany and KSB SE & Co. KGaA, Germany), Sandro Schulze (Anhalt University of Applied Sciences, Germany), and Jacob Krüger (Eindhoven University of Technology, The Netherlands)

ACER: An AST-Based Call Graph Generator Framework 254
Andrew Chen (College of William and Mary, USA), Yanfu Yan (College of William and Mary, USA), and Denys Poshyvanyk (College of William and Mary, USA)

Reproducing and Improving the BugsInPy Dataset 260
Faustino Aguilar (University of Panama, Panama), Samuel Grayson (University of Illinois Urbana-Champaign, USA), and Darko Marinov (University of Illinois Urbana-Champaign, USA)

New Ideas and Emerging Results Track

Can We Trust the Default Vulnerabilities Severity? 265
Matteo Esposito (University of Rome Tor Vergata), Sergio Moreschini (Tampere University), Valentina Lenarduzzi (University of Oulu), David Hästbacka (Tampere University), and Davide Falessi (University of Rome Tor Vergata)

On the Impact of Language Selection for Training and Evaluating Programming Language Models 271
Jonathan Katzy (Delft University of Technology, The Netherlands), Maliheh Izadi (Delft University of Technology, The Netherlands), and Arie van Deursen (Delft University of Technology, The Netherlands)

Uncovering the Hidden Risks: The Importance of Predicting Bugginess in Untouched Methods 277
Matteo Esposito (University of Rome Tor Vergata) and Davide Falessi (University of Rome Tor Vergata)

Author Index **283**