2023 IEEE/ACM 2nd International Workshop on Natural Language-Based Software Engineering (NLBSE 2023)

Melbourne, Australia
20 May 2023

IEEE

IEEE Catalog Number: CFP23CD3-POD
ISBN: 979-8-3503-0179-3

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:	CFP23CD3-POD
ISBN (Print-On-Demand):	979-8-3503-0179-3
ISBN (Online):	979-8-3503-0178-6

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/ACM 2nd International Workshop on Natural Language-Based Software Engineering (NLBSE)

NLBSE 2023

Table of Contents

Message from the Program Chairs ..... **vii**
Committees ..... **ix**

2023 IEEE/ACM 2nd International Workshop on Natural Language-Based Software Engineering (NLBSE)

The NLBSE'23 Tool Competition 1
Rafael Kallis (Rafael Kallis Consulting), Maliheh Izadi (Delft University of Technology), Luca Pascarella (ETH Zurich), Oscar Chaparro (College of William & Mary), and Pooja Rani (University of Zurich)

The (ab)use of Open Source Code to Train Large Language Models 9
Ali Al-Kaswan (Delft University of Technology, The Netherlands) and Maliheh Izadi (Delft University of Technology, The Netherlands)

Generalizability of NLP-based Models for Modern Software Development Cross-Domain Environments 11
Rrezarta Krasniqi (University of North Texas) and Hyunsook Do (University of North Texas)

An Intelligent Tool for Classifying Issue Reports 13
Muhammad Laiq (Blekinge Institute of Technology)

Few-Shot Learning for Issue Report Classification 16
Giuseppe Colavito (University of Bari, Italy), Filippo Lanubile (University of Bari, Italy), and Nicole Novielli (University of Bari, Italy)

Performance Comparison of Binary Machine Learning Classifiers in Identifying Code Comment Types: An Exploratory Study 20
Amila Indika (University of Hawaii at Mānoa), Peter Washington (University of Hawaii at Mānoa), and Anthony Peruma (University of Hawaii at Mānoa)

Classifying Code Comments via Pre-trained Programming Language Model 24
Ying Li (Southern University of Science and Technology), Haibo Wang (Southern University of Science and Technology), Huaien Zhang (Southern University of Science and Technology, The Hong Kong Polytechnic University), and Shin Hwei Tan (Southern University of Science and Technology)

STACC: Code Comment Classification using SentenceTransformers 28
Ali Al-Kaswan (Delft University of Technology, The Netherlands), Maliheh Izadi (Delft University of Technology, The Netherlands), and Arie van Deursen (Delft University of Technology, The Netherlands)

An Exploratory Study on the Usage and Readability of Messages Within Assertion Methods of Test Cases 32
Taryn Takebayashi (University of Hawaii at Mānoa), Anthony Peruma (University of Hawaii at Mānoa), Mohamed Mkaouer (Rochester Institute of Technology), and Christian Newman (Rochester Institute of Technology)

Stop Words for Processing Software Engineering Documents: Do they Matter? 40
Yaohou Fan (The University of Melbourne, Australia), Chetan Arora (Monash University, Australia), and Christoph Treude (The University of Melbourne, Australia)

Applying information theory to software evolution 48
Adriano Torres (University of Adelaide), Sebastian Baltes (University of Adelaide), Christoph Treude (University of Melbourne), and Markus Wagner (Monash University)

Zero-shot Prompting for Code Complexity Prediction Using GitHub Copilot 56
Mohammed Latif Siddiq (University of Notre Dame, USA), Abdus Samee (Bangladesh University of Engineering and Technology, Bangladesh), Sk Ruhul Azgor (Bangladesh University of Engineering and Technology, Bangladesh), Md. Asif Haider (Bangladesh University of Engineering and Technology, Bangladesh), Shehabul Islam Sawraz (Bangladesh University of Engineering and Technology, Bangladesh), and Joanna C. S. Santos (University of Notre Dame, USA)

Evaluating Code Comment Generation With Summarized API Docs 60
Bilel Matmti (University of British Columbia) and Fatemeh Fard (University of British Columbia)

Author Index **65**