## IJARSCT



International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Volume 5, Issue 3, June 2025



## Semantically Enhanced AST Vectorization for Scalable Code Clone Detection

Kakumanu Kailash Nath, Cheemaladinne Lakshmi Siri, Gugulothu Karthik

B.Tech Students, Dept. of Computer Science and Engineering

Dr. R. Lakshmi Tulasi

Professor, Dept. of Computer Science and Engineering R.V.R & J.C College of Engineering, Guntur, Andhra Pradesh, India

Abstract: Code clone detection is a critical task in software engineering, aimed at identifying duplicated code segments that can hinder maintainability and increase the risk of defect propagation. Code2Img presents a scalable solution by converting Abstract Syntax Trees (ASTs) into image-based vector representations, enabling effective detection of syntactically complex clones. However, its reliance on structural similarity limits its ability to capture deeper semantic relationships. In this work, we enhance Code2Img by integrating a semantic transition scoring mechanism. We compute heuristic scores for AST node transitions based on node depth and connectivity within local function contexts, reflecting their semantic roles. These scores are further weighted using Inverse Document Frequency (IDF) across a broader corpus to emphasize informative yet uncommon transition patterns. The enriched semantic scores are incorporated into the vector representation used by Code2Img, enhancing its capacity to detect nuanced Type-3 clones and extending potential coverage toward Type-4 clones. Empirical evaluation demonstrates that our approach preserves the scalability of the original framework while significantly improving semantic clone detection performance.

Keywords: Code clone

Copyright to IJARSCT www.ijarsct.co.in



