I-SEE

Integrated Software Ecosystem Evaluation

Programm / Ausschreibung	KS 24/26, KS 24/26, BRIDGE 2025/01	**Status**	laufend
Projektstart	01.02.2026	**Projektende**	31.07.2028
Zeitraum	2026 - 2028	**Projektlaufzeit**	30 Monate
Projektförderung	€ 356.802		
Keywords	assembly theory; software security; software complexity; software analysis; fault prediction		

Projektbeschreibung

Softwaresysteme gehören zu den komplexesten menschengemachten Artefakten, mit Millionen von Codezeilen, tausenden Abhängigkeiten und einer kontinuierlichen Weiterentwicklung durch verteilte Teams. Diese Komplexität führt zu Risiken für Qualität und Sicherheit, mit erheblichen wirtschaftlichen und gesellschaftlichen Auswirkungen. Defekte und Schwachstellen entstehen selten zufällig; sie gehen aus erkennbaren Entstehungsmustern hervor. Diese Muster werden durch Metriken wie Code-Churn, strukturelle Komplexität oder fragmentierte Zuständigkeiten abgebildet, da solche Metriken die zugrunde liegenden sozio-technischen und evolutionären Faktoren erfassen, die Systeme fragiler machen.
Bestehende Ansätze behandeln Qualität und Sicherheit weitgehend getrennt und nutzen damit nicht die Möglichkeit, frühe Qualitätsindikatoren auch als Prädiktoren für Sicherheitsrisiken einzusetzen. Notwendig ist ein einheitliches Rahmenwerk, das erklärt, wie Schwachstellen aus der Softwareevolution hervorgehen, und interpretierbare Signale für frühzeitige Eingriffe liefert.
Dieses Projekt entwickelt ein solches Rahmenwerk durch die Anwendung der Assembly Theory, die ursprünglich in den Naturwissenschaften entwickelt wurde, auf Softwaresysteme. Assembly Theory beschreibt Artefakte als Ergebnisse schrittweiser Entstehungspfade, messbar über Indizes wie Pfadlänge, Diversität und Komplexität. Übertragen auf Software bedeutet dies, Code, Abhängigkeiten und Entwicklungshistorien als Assembly-Pfade darzustellen und Risiken mit ihrer strukturellen Entstehung zu verknüpfen.
Das Projekt verfolgt drei Ziele:

Die Formalisierung der Assembly Theory für Softwaresysteme und die Definition berechenbarer Metriken für Entstehungspfade.

Die Integration dieser Metriken in Vorhersagemodelle für Defekte und Schwachstellen, wobei Defektvorhersage als Grundlage für Schwachstellenbewertung dient.

Die Validierung des Ansatzes anhand von Open-Source- und proprietären Repositorien aus Industrieprojekten, ergänzt durch

Visualisierungs- und Interpretierbarkeitswerkzeuge für den praktischen Einsatz.

Die erwarteten Ergebnisse umfassen ein prototypisches Werkzeug für prädiktive Wartung und Sicherheitsanalysen, eine theoretische Grundlage für pfadbasierte Softwareanalytik sowie empirische Evidenz, dass Assembly-Theory-basierte Metriken frühere und klarer interpretierbare Risikoindikatoren liefern als herkömmliche Ansätze.

Abstract

Software systems are among the most complex human-made artifacts, with millions of lines of code, thousands of dependencies, and continuous evolution by distributed teams. This complexity creates risks for both quality and security, with major economic and societal impact. Defects and vulnerabilities rarely appear spontaneously; they emerge from recognizable formation patterns. These patterns are depicted by metrics such as code churn, structural complexity, or fragmented ownership, since such metrics capture the underlying socio-technical and evolutionary factors that increase fragility.
Existing approaches largely treat quality and security separately, missing the opportunity to use early quality indicators as predictors of security risk. What is needed is a unified framework that explains how vulnerabilities arise from software evolution and provides interpretable signals for early intervention.
This project introduces such a framework by applying Assembly Theory, originally developed in the natural sciences, to software systems. Assembly Theory models artifacts as outcomes of stepwise formation pathways, measured by indices such as pathway length, diversity, and complexity. Applied to software, this means representing code, dependencies, and development histories as assembly pathways and linking risk to their structural formation, making risk indicators more explanatory and transferable.
The project pursues three goals:

Formalize Assembly Theory for software systems and define computable measures of formation pathways.

Integrate these measures into predictive models for defects and vulnerabilities, using defect prediction as a basis for vulnerability assessment.

Validate the approach on open-source and proprietary industrial repositories, with visualization and interpretability tools for practical use.

Outcomes include a prototype tool for predictive maintenance and vulnerability analytics, a theoretical foundation for pathway-based software analysis, and empirical evidence that assembly-derived metrics provide earlier and more interpretable risk indicators than conventional approaches.

Projektkoordinator

- SBA Research gemeinnützige GmbH

Projektpartner

- CONDIGNUM GmbH