

科学研究費助成事業（基盤研究（S））公表用資料
〔令和4（2022）年度 中間評価用〕

令和4年3月31日現在

研究期間：2020年度～2024年度
課題番号：20H05706
研究課題名：次世代ソフトウェアエコシステムのための基盤・展開技術

研究代表者氏名（ローマ字）：松本 健一（MATSUMOTO Kenichi）
所属研究機関・部局・職：奈良先端科学技術大学院大学・先端科学技術研究科・教授
研究者番号：70219492

研究の概要：

本研究課題では、AI、自然言語処理、サービス化、ブロックチェーン等の最新デジタル技術を積極的に活用、連携させることで、ソフトウェア開発運用の今日的な技術的課題を解決し、これからのソフトウェアエコシステムを支える技術へと展開する。ソフトウェア開発運用に係る無駄を極力排除することで高い経済性と持続可能性を実現し、より一層の技術革新と新たな技術体系の構築につながる成果を目指す。

研究分野：ソフトウェア工学

キーワード：ソフトウェア再利用、マイクロサービス、AI技術応用、ブロックチェーン技術

1. 研究開始当初の背景

IoT や AI 等の新たなデジタル技術の進展は、ソフトウェアの重要性と多様性をますます高めている。ソフトウェア開発運用が直面する技術的課題を解決し、より高い品質のソフトウェアを社会全体で蓄積、共有、循環するための、より一層の技術革新が求められている。本研究課題では、「技術的負債の増加」、「人的資源の不足」、および、「技術情報の多様化」を喫緊に解決すべき技術的課題と位置づける。

2. 研究の目的

本研究課題では、ソフトウェア開発運用に欠かせない資源を「プロダクト」、「人」、および、「技術情報」という3つの観点で捉え、AI、自然言語処理、サービス化、ブロックチェーン等の最新デジタル技術を積極的に活用、連携させることで、開発運用の今日的な技術的課題を解決し、これからのソフトウェアエコシステムを支える技術へと展開する。ただし、ソフトウェア開発運用の完全自動化を目指すものではない。あくまでも人間中心の社会を想定し、ソフトウェア開発運用に係る無駄を極力排除することで高い経済性と持続可能性を実現することを目的とする。

3. 研究の方法

ソフトウェア開発運用の技術的課題を次に示す3つの「学術的問い（Research Question）」として明確に示し、「アップサイクリング」、「適格性」、「外部技術情報とのリンク」といった独自の着想・アプローチに基づく研究開発でその答えに迫る。

RQ1:プロダクト、特に、プログラムコードのアップサイクリングは可能か？

プログラムコードに内在する技術的負債の様態解明を通じて、負債とされるコードを価値あるソフトウェア資産（マイクロサービス等）に転換（アップサイクリング）することは可能か？

RQ2:人とAIを適材適所に配置することで人的資源の不足解消は可能か？

ソフトウェア開発運用における人的資源の様態解明を通じて、開発運用における人とAIの適格性評価（Competency Evaluation）を実現し、開発運用の生産性向上と開発運用者のスキル転換を促進することは可能か？

RQ3:外部技術情報ともリンクする堅牢で持続可能な品質管理は可能か？

ナレッジコミュニティや学術論文誌等から得られる外部技術情報を、ソフトウェアプロダクトと紐づけ、改ざんが不可能で第三者検証も可能な形式で保持することで、技術的負債を抑止・低減することは可能か？

研究開発項目	予備調査・研究開発実績等	独自技術・アプローチ	既存技術・アプローチ
プロダクトのアップサイクリング	プログラムコードからのビジネスロジック・イデオムの抽出（金融・保険業向けアプリケーションでは、18イデオムでプログラムコードの80%以上を網羅。）	コードクローン解析 レガシーソフトウェアからのサービス抽出 品質属性に基づくAPI価値考察 ソフトウェアライブラリ間の依存度評価	データフロー解析 SATD (Self-Admitted Technical Debt)
人とAIの適材適所配置	開発運用者の活動履歴に基づく分類・貢献様態の類型化 OSSコミュニティのヒューマン・キャピタル評価・構成員動態予測	貢献度等に基づく適格性評価 スキル指標とベンチマーキング ソフトウェアエコシステムのモデル化・可視化	クラスタリング、原型分析、位相的データ解析 人の分散表現 (User Embeddings)
外部技術情報ともリンクする品質管理	プログラムコードのコメント中の学術論文参照箇所の特定、URL調査	ブロックチェーンを用いたソフトウェアビルドプロセス記録・監視法（特許第6692013号）	固有表現抽出 分散デジタルID

機械学習
自然言語
処理

4. これまでの成果

(1) プロダクトのアップサイクリング：技術的負債の解析，コード要約

「技術的負債」と「コード要約」という二つの観点で、プロダクト中のアップサイクル素材を特定、抽出する技術を開発した。更に、アップサイクル素材のカタログ化、アップサイクル知識ベース、コンテキストウェア・サービス連携基盤等、マイクロサービス化によるプログラムコードの資産化につながる基盤技術の開発を進めた。

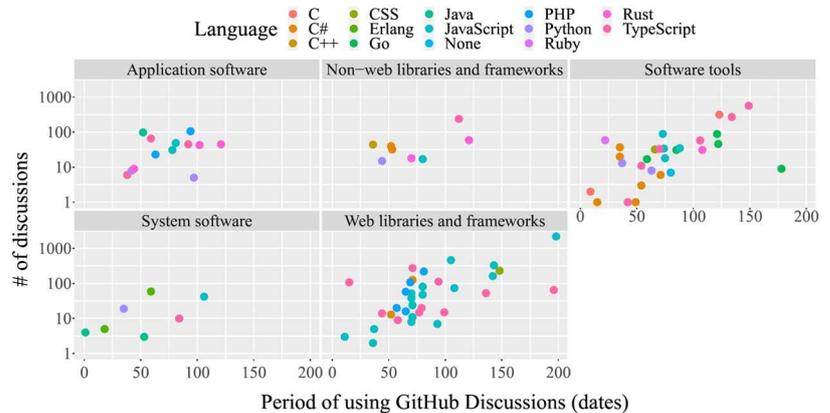
(2) 人と AI の適材適所配置：人的資源の解析、スキル定量化

ソフトウェアライブラリの実績に基づいて開発運用者のスキルを分析する手法を開発した。また、ヨーロッパ言語共通参照枠 CEFR を規範として、プログラミング言語 Python を用いたプログラミングにおける適格性(スキル習熟度レベル)を評価するフレームワークを開発した。

(3) 外部技術情報ともリンクする

品質管理：外部技術情報の解析、プロダクトとの紐づけ

ソフトウェアに内在する技術的負債に関連するバグレポートを特定する技術を開発し、情報を集約、提示するシステムを **GitHub Bot** として実装した。また、ソフトウェア開発運用に係るナレッジコミュニティにおける議論展開を解析し、外部技術情報ともリンクする品質管理に資する知見を得た。



以上の成果の多くは、膨大な研究資料(ソフトウェア開発運用履歴データ)を集積、解析することで得られている。集積規模は既にソフトウェア開発運用プロジェクト1000万件に達している。

5. 今後の計画

これまでに得られた研究成果を基盤として、令和4年、5年度には「展開技術の開発」を、最終年度となる令和6年度には、「基盤・展開技術の実証実験」をそれぞれ実施する。ただし、新型コロナウイルス感染症の感染拡大はいずれ終息するとしても、本研究課題をとりまく環境が感染拡大以前の状態に戻るとは限らない。次の点に留意し、研究目的の着実な達成を目指す。

- (1) 膨大な研究資料の活用・更新
- (2) オンラインツール・システムのこれまで以上の導入・活用、バーチャルラボの構築
- (3) 独自開催する国内・国際研究集会の強化、オンライン・オンサイトのシームレス化

6. これまでの発表論文等(受賞等も含む)

- [1] Takuya Nakata, Sinan Chen, and Masahide Nakamura, “Uni-Messe: Unified Rule-Based Message Delivery Service for Efficient Context-Aware Service Integration,” *Energies*, Vol.15, No.5, 1729, February 2022.
- [2] Hideaki Hata, Nicole Novielli, Sebastian Baltes, Raula Gaikovina Kula, and Christoph Treude, “GitHub Discussions: An Exploratory Study of Early Adoption,” *Empirical Software Engineering*, Vol.27, Issue 1, Article No.3, January 2022.
- [3] Keita Takenouchi, Takashi Ishio, Joji Okada, and Yuji Sakata, “PATSQL: Efficient Synthesis of SQL Queries from Example Tables with Quick Inference of Projected Columns,” *Proceedings of the VLDB Endowment*, Vol.14, No.11, pp.1937-1949, July 2021.
- [4] Gregorio Robles, Raula Gaikovina Kula, Chaiyong Ragkhitwetsagul, Tattiya Sakulniwat, Kenichi Matsumoto, and Jesus M. Gonzalez-Barahona, “pycefr: Python Competency Level through Code Analysis,” In Proc. of 30th IEEE/ACM International Conference on Program Comprehension (ICPC 2022) Tool Track (to appear).
- [5] Wilson Chukwu Emmanuel, and Akito Monden, “Human Resource Analysis Based on Used Libraries in Eclipse Projects on GitHub,” In Proc. of 22nd IEEE-ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel Distributed Computing (SNPD 2021), pp.254-257, November 2021.

【以上、すべて「査読あり」の学術論文誌論文・国際会議発表】

7. ホームページ等

ホームページ：<https://naist-se.github.io/>

バーチャルラボ：<https://selab-naist.ovice.in/>