

受入大学名	奈良先端科学技術大学院大学		
Host University	Nara Institute of Science and Technology		
外国人研究者	パンナチッタ パサコーン		
Foreign Researcher	PHANNACHITTA Passakorn		
受入研究者	松本健一	職名	教授
Research Advisor	MATSUMOTO Kenichi	Position	Professor
受入学部/研究科	先端科学技術研究科		
Faculty/Department	Graduate School of Science and Technology		

<外国人研究者プロフィール/Profile>

国籍	タイ
Nationality	Thai
所属機関	チェンマイ大学 芸術・メディア・技術学部
Affiliation	College of Arts, Media and Technology, Chiang Mai University, Thailand
現在の職名	助教
Position	Assistant Professor
研究期間	2022年12月30日 ~ 2023年3月18日 (78日間)
Period of Stay	78 days (Dec 30, 2022 - Mar 18, 2023)
専攻分野	ソフトウェア工学
Major Field	Software engineering



受入研究室にて/ In the host lab.

<外国人研究者からの報告/Foreign Researcher Report>

①研究課題 / Theme of Research
Mining software repositories attempt to acquire evidence-based insights from software archives. These insights have assisted software development activities, such as reducing software development costs and increasing software quality. More recently, the software archives most valuable are textual-based archives written in natural language. They include source code comments, defect tickets, and many discussion log messages throughout the software development lifecycle. Typically, extracting actionable insight from these textual data is done by applying information-retrieval techniques over these texts. These analyses require high-quality data input for generating the most accurate and comprehensible insight possible.
②研究概要 / Outline of Research
It is observed that a massive appearance of typos is typical in these software archives, and these massive typos can drastically degrade the quality of textual analysis results or lead to incorrect conclusions. However, there is no attempt to study the statistics and effects of the appearance of these massive typos in mining software repositories researches and practices. Fixing these typos is not straightforward mainly because texts appearing on online platforms are typically casual, and typos are more forgiven than other communication mediums. This research explores the possibility of adopting a syllabification-based spell corrector to correct misspelled textual data in software archives.
③研究成果 / Results of Research
Although the proposal initially focused on the distribution of typos, further analysis revealed that it did not strongly impact the quality of the analytical process. Therefore, the decision was made to halt this direction of the study and explore other possibilities. During the visit, the researcher was able to participate as a co-author in an article submitted to an ACM conference on Innovation and Technology in Computer Science Education (ITiCSE) to work is to better understand how programmers learn about programming. Further, the researcher received invaluable advice from the professors he visited, which inspired him to pursue their passion in research and collaboration in the direction that brings meaningful contributions to society.
④今後の計画 / Further Research Plan
Upon returning to Thailand, the researcher is excited to acquire data from the real sector and leverage his primary skills and knowledge in areas such as software engineering, AI, data mining, and information retrieval to provide vital support to the industry. This realization was made during multiple business trips during this visit. The researcher believes that the achieved research outcome will both publishable and impactful in a developing country like Thailand. The researcher's primary takeaway from this visit is that the level of acceptance from society is the best quality metric. By focusing on creating real-world impact and seeking the acceptance of the Thai society, the researcher hopes to make a meaningful contribution to the development of his country.

<受入研究者からの報告/Research Advisor Report>

①研究課題 / Theme of Research

本研究では、ソフトウェア開発で用いられる設計書やバグ報告書など各種ドキュメントに含まれる自然言語文における誤記（スペルミス等）を、シラビフィケーション（分綴法）に基づき検出、校正する手法の妥当性を検証した。ソフトウェア開発で用いられるそうしたドキュメントには、ソフトウェア開発のコスト削減やソフトウェアの品質向上に資する情報が含まれているとされているが、同時に、誤記が多数含まれていることでも知られている。本研究では、シラビフィケーションにより誤記を検出し、校正することがどこまで可能か、また、その検出と校正がソフトウェアの開発コストや品質に与える影響について、ソフトウェアリポジトリマイニング分野で国際的な研究成果を挙げてきている我が国の研究者との討議や意見交換を通じて明らかにした。

②研究指導概要 / Outline of Research

受入研究者が構築しているオープンソースソフトウェア開発履歴データリポジトリに格納されているデータを対象に、外国人研究者が提案する誤記検出・構成手法を適用した。当該リポジトリには、オープンソースソフトウェアを開発する約1,000万プロジェクトのデータが集積されている。ただし、提案手法が対象とするドキュメント類が作成されていないプロジェクトなどが含まれていることから、提案手法の妥当性評価の目的と条件に適合するプロジェクト群を特定したうえで提案手法を適用し、分析を行った。適用・分析対象とした開発ドキュメントは、ソースコード（に含まれるコメント）、バグ票、開発者間でやり取りされたメッセージログ等である。

③研究指導成果 / Results of Research

提案手法をオープンソースソフトウェア開発履歴データに適用し、開発時に作成されたドキュメントに含まれる誤字の分布などを詳細に分析した結果、ソフトウェアの開発コストや品質との間に顕著な関係性を発見することはできなかった。ただし、ドキュメントに含まれる誤記がそのドキュメントを特徴づけることには変わりなく、別の角度からの分析、そして、ソフトウェア開発支援に結びつけるアプローチを検討することとした。なお、研究を進める中で、プログラミングの学習プロセスに関する議論を行い、その成果をまとめた論文を国際会議ACM Innovation and Technology in Computer Science Education (ITiCSE) に外国人研究者との共著論文として投稿することができた。また、我が国の研究者との討議や意見交換を通じて、外国人研究者が、研究への情熱と社会貢献に向けたコラボレーションへのモチベーションを高めたことも、研究指導の成果の一つと言える。

④留学生交流事業の活動状況 / Activities of International Student Exchange Program

当初の計画通り、岡山大学 自然科学学域 門田暁人教授、および、神戸大学 数理・データサイエンスセンター 中村匡秀副センター長・教授を訪問し、討議や意見交換を行った。また、受入研究者の所属機関である奈良先端科学技術大学院大学において、博士前期課程学生を対象とした科目「コロキウムA/B」での講演、また、受入研究者が主催する「次世代ソフトウェアエコシステムワークショップ」での特別講演を行い、学生や若手研究者との討議や意見交換も行った。

⑤今後の計画 / Further Research Plan

外国人研究者は、母国タイに帰国後、ソフトウェア開発履歴データを企業等から取得し、ソフトウェア工学、AI、データマイニング、情報検索といった分野において、今回の来日において得られた知見を活用し、ソフトウェア開発に不可欠な支援技術の研究開発に取り組む予定である。研究成果が社会に受け入れられるかどうか、現実社会にどれだけインパクトを与えることができるか、母国タイのような新興国では、特に重要な評価指標になると考えられる。そのような研究成果を生み出すことで、母国タイの発展に貢献したい。受入大学だけでなく、岡山大学や神戸大学を訪れ、当該分野で国際的にも活躍する研究者との交流の中で外国人研究者に芽生えた強い思いである。



受入大学における講演 / Lecture at the host university



神戸大学を訪問 / Visiting Kobe University