Linking Software System Artifacts: Toward Generic Traceability Link Recovery through Retrieval-Augmented Generation

Dominik Fuchß [1], Tobias Hey [1], Jan Keim [1], Haoyu Liu [1], Niklas Ewald [1], Tobias Thirolf [1], and Anne Koziolek [1]

Abstract: Our paper [Fu25a], published at the 47th IEEE International Conference on Software Engineering (ICSE), introduces a generic approach to traceability link recovery (TLR) leveraging Retrieval-Augmented Generation (RAG).

Keywords: Traceability Link Recovery, Large Language Models, Retrieval-Augmented Generation

Introduction The development and maintenance of software systems require managing numerous interrelated artifacts. Understanding these complex interconnections is essential for many software engineering tasks. Existing automated TLR methods predominantly focus on linking specific artifact types, such as requirements and code. However, recent advances in large language models (LLMs) offer the potential to enable TLR approaches with broader applicability.

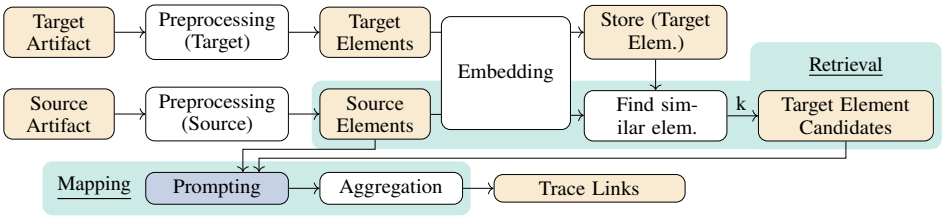

Fig. 1: Overview of the LiSSA approach [Fu25a]. Data is represented in orange, prompting is shown in blue, and other processing is displayed in white.

Methods The paper presents Linking Software System Artifacts (LiSSA), a framework that harnesses the capabilities of LLMs and enhances them through RAG. The framework allows generic TLR, meaning that it is not restricted to one specific TLR task.

[1] Karlsruhe Institute of Technology, Germany,
dominik.fuchss@kit.edu, https://orcid.org/0000-0001-6410-6769;
hey@kit.edu, https://orcid.org/0000-0003-0381-1020;
jan.keim@kit.edu, https://orcid.org/0000-0002-8899-7081;
haoyu.liu@kit.edu, https://orcid.org/0009-0002-7676-5010;
niklas.ewald@alumni.kit.edu, https://orcid.org/0009-0000-8868-0562;
tobias.thirolf@student.kit.edu, https://orcid.org/0009-0006-7052-4020;
koziolek@kit.edu, https://orcid.org/0000-0002-1593-3394

doi:10.18420/se2026-21

Research Questions

RQ1: Is a generic RAG-based TLR approach better than task-specific, state-of-the-art approaches?

RQ2: Is chain-of-thought prompting more effective than a simple classification prompt?

RQ3: Does retrieval and mapping on a fine-grained, sub-artifact level improve the TLR performance compared to mapping on an artifact level?

RQ4: Does RAG improve the TLR performance compared to embedding-based information-retrieval TLR?

Results We empirically evaluate LiSSA on three TLR tasks: requirements-to-code, documentation-to-code, and architecture documentation-to-architecture models. We compared our approach against state-of-the-art baselines. The results demonstrate that the RAG-based method significantly outperforms existing approaches on code-related tasks. We also find that chain-of-thought prompting outperforms simple classification prompts in F_1-score. Examining the various preprocessing techniques and the effect of fine-grained mappings, we demonstrate that, on average, preprocessing artifacts is not beneficial. Finally, we find that, on average, an LLM-based classification outperforms retrieval-only in all considered TLR tasks for both F_1-score and F_2-score.

Conclusion This research advances the field of traceability by introducing a novel RAG-based approach for TLR. We provide insights into the impact of different prompt types and preprocessing techniques on performance. Our findings open up new research directions for the application of LLMs to TLR tasks. To promote replicability, transparency, and extensibility, we release the LiSSA source code, datasets, and results as part of our replication package [Fu25b], enabling further research and development in this area.

References

[Fu25a] Fuchß, D. et al.: LiSSA: Toward Generic Traceability Link Recovery through Retrieval-Augmented Generation. In: Proceedings of the IEEE/ACM 47th International Conference on Software Engineering. ICSE '25, Institute of Electrical and Electronics Engineers (IEEE), Ottawa, Canada, 2025.

[Fu25b] Fuchß, D. et al.: Replication Package: LiSSA: Toward Generic Traceability Link Recovery through Retrieval-Augmented Generation, https://doi.org/10.5281/zenodo.14714706, 2025.