



## Recovering Trace Links Between Software Documentation And Code

Jan Keim <sup>1</sup>, Sophie Corallo <sup>1</sup>, Dominik Fuchß <sup>1</sup>, Tobias Hey <sup>1</sup>, Tobias Telge <sup>1</sup>, and Anne Koziolk <sup>1</sup>

**Keywords:** software traceability, software architecture, documentation, transitive links, intermediate artifacts, information retrieval

Our paper [Ke24] that was published at the 2024 46th IEEE International Conference on Software Engineering (ICSE) is about Traceability Link Recovery (TLR) between Software Architecture Documentations and Code using Software Architecture Models as intermediate artifacts.

**Introduction** Software development involves creating various artifacts at different levels of abstraction and establishing relationships between them is essential. traceability link recovery (TLR) automates this process, enhancing software quality by aiding tasks like maintenance and evolution. However, automating TLR is challenging due to semantic gaps resulting from different levels of abstraction. While automated TLR approaches exist for requirements and code, architecture documentation lacks tailored solutions, hindering the preservation of architecture knowledge and design decisions.

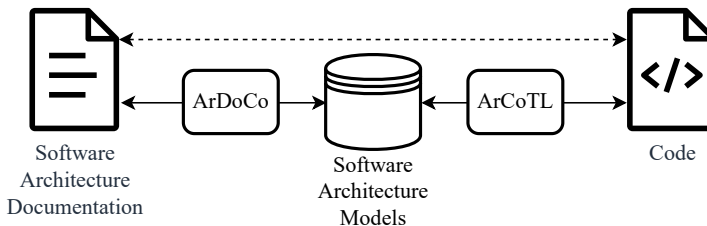








Fig. 1: High-level view of the Transitive links for Architecture and Code (TransArC) approach for linking software architecture documentations (SADs) and code using Architecture Documentation Consistency (ArDoCo) [Ke23b] and our novel ARchitecture-to-CODE Trace Linking (ArCoTL).

<sup>1</sup> Karlsruhe Institute of Technology, Germany, jan.keim@kit.edu,  <https://orcid.org/0000-0002-8899-7081>;  
sophie.corallo@kit.edu,  <https://orcid.org/0000-0002-1531-2977>;  
dominik.fuchss@kit.edu,  <https://orcid.org/0000-0001-6410-6769>;  
hey@kit.edu,  <https://orcid.org/0000-0003-0381-1020>;  
tobias.telge@alumni.kit.edu,  <https://orcid.org/0009-0002-6700-6426>;  
koziolk@kit.edu,  <https://orcid.org/0000-0002-1593-3394>

**Methods** The paper presents our approach TransArC for TLR between architecture documentation and code, using component-based architecture models as intermediate artifacts to bridge the semantic gap. The idea is to combine two specialized approaches to improve the results for the wider semantic gap between the original artifacts. We create transitive trace links by combining the existing approach ArDoCo for linking architecture documentation to models with our novel approach ArCoTL for linking architecture models to code.

## Research Questions

- RQ1** How well can our approach ArCoTL recover trace links between component-based software architecture models and code?
- RQ2** How accurate can our approach TransArC using intermediate artifacts recover trace links between software architecture documentation and code?
- RQ3** How do the results for linking software architecture documentation and code compare to state-of-the-art requirements-to-code approaches?

**Results** We evaluate our approaches with five open-source projects, comparing our results to baseline approaches. The model-to-code TLR approach achieves an average  $F_1$ -score of 0.98, while the documentation-to-code TLR approach achieves a promising average  $F_1$ -score of 0.82, significantly outperforming baselines.

**Conclusion** Combining two specialized approaches with an intermediate artifact shows promise for bridging the semantic gap. To ensure replicability and transparency, we have made available a comprehensive replication package [Ke23a], encompassing the implemented approach, baseline models, evaluation data, and the obtained results. In future research, we plan to explore further possibilities for such transitive approaches.

## Bibliography

- [Ke23a] Keim, Jan; Corallo, Sophie; Fuchß, Dominik; Hey, Tobias; Telge, Tobias; Koziolk, Anne: Replication Package for "Recovering Trace Links Between Software Documentation And Code". Zenodo, 2023.
- [Ke23b] Keim, Jan; Corallo, Sophie; Fuchß, Dominik; Koziolk, Anne: Detecting Inconsistencies in Software Architecture Documentation Using Traceability Link Recovery. In: 2023 IEEE 20th International Conference on Software Architecture (ICSA). pp. 141–152, 2023.
- [Ke24] Keim, Jan; Corallo, Sophie; Fuchß, Dominik; Hey, Tobias; Telge, Tobias; Koziolk, Anne: Recovering Trace Links Between Software Documentation And Code. In: Proceedings of 46th IEEE International Conference on Software Engineering (ICSE 2024). 2024.