

Benchmark for Graph Schema Evolution (Bachelors Thesis)

Topic Description To modify a system in alignment with changing requirements such as newly emerging user needs, schema evolution is necessary. Schema evolution is executed via so-called evolution operations, namely `add`, `rename`, `delete`, `copy`, `move`, `split`, `merge` and the graph-specific `transform` operation [4]. When testing a tool that executes schema evolution (like NAUTILUS[3] for graph schema evolution) a adequate dataset needs to be used, whose characteristics will be analyzed in this thesis. The topic focuses on graph databases, excelling in the storage of highly interconnected data like it is the case for social networks.

Tasks This thesis aims to identify the characteristics of a dataset to analyze schema evolution. This includes

- (1) the identification of characteristics needed in a dataset to analyze schema evolution (here: `add`, `rename`, `delete` and `transform`),
- (2) the selection and generation of aligning graph datasets,
- (3) the creation and selection of different workloads, i.e., sets of conducted evolution operations,
- (4) and an implementation with Python and Neo4j.

Furthermore, a comparison of the different workloads and datasets is expected to show the differences as well as advantages and disadvantages in each case. This means, that use case specifics are thematized during the comparison. For testing the state-of-the-art graph database Neo4j will be used.

General Information

- Contact: Dominique Hausler (dominique.hausler@ur.de)
- Prerequisites:
 - Programming skills: Python
 - Successfully completed lecture in *Datenbanken I* (Beneficial but not mandatory: knowledge about graph databases)

Start Literature

- Angela Bonifati et al. “Schema Validation and Evolution for Graph Databases”. In: *ER*. vol. 11788. Lecture Notes in Computer Science. Springer, 2019, pp. 448–456
- Renzo Angles et al. “PG-Schema: Schemas for Property Graphs”. In: *Proc. ACM Manag. Data* 1.2 (2023), 198:1–198:25

References

- [3] Dominique Hausler and Meike Klettke. “Nautilus: Implementation of an Evolution Approach for Graph Databases”. In: *MoDELS (Companion)*. ACM, 2024, pp. 11–15.
- [4] Dominique Hausler, Meike Klettke, and Uta Störl. “A language for graph database evolution and its implementation in Neo4j”. In: *ER (Companion)*. Vol. 3618. CEUR Workshop Proceedings. CEUR-WS.org, 2023.