

Graph Schema Matching and Mapping (Bachelors Thesis)

Topic Description To detect schema changes between two different versions or databases, their schemas need to be compared with each other. Therefore that graph databases allow directly inserting data without pre-defining a schema, the schema first needs to be extracted. After extracting both schemas, they are compared by detecting different and identical structures. Followed, by a detection and categorization of the differences. Schema-less databases hereby hold additional challenges compared to relational databases, namely syntactical errors like typos or semantic errors like synonyms. Consequently, a relaxed matching and mapping approach is mandatory.

Tasks This thesis aims to

- (1) compare known algorithms when matching and mapping schemas,
- (2) design an algorithm (based on the literature + own extensions) to identify changes,
- (3) and implement and evaluate the own matching and mapping algorithm, incorporating specifics of graph databases (e.g., how to handle possibly occurring typos).

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

- Erhard Rahm and Philip A. Bernstein. “A survey of approaches to automatic schema matching”. In: *VLDB J.* 10.4 (2001), pp. 334–350
- Gwendal Daniel, Gerson Sunyé, and Jordi Cabot. “UMLtoGraphDB: Mapping Conceptual Schemas to Graph Databases”. In: *ER*. vol. 9974. Lecture Notes in Computer Science. 2016, pp. 430–444
- Leonard Wörteler et al. “Cardinality Estimation using Label Probability Propagation for Subgraph Matching in Property Graph Databases”. In: *EDBT*. OpenProceedings.org, 2022, 2:285–2:297
- Fausto Giunchiglia and Daqian Shi. “Property-based entity type graph matching”. In: *OM@ISWC*. vol. 3063. CEUR Workshop Proceedings. CEUR-WS.org, 2021, pp. 25–36