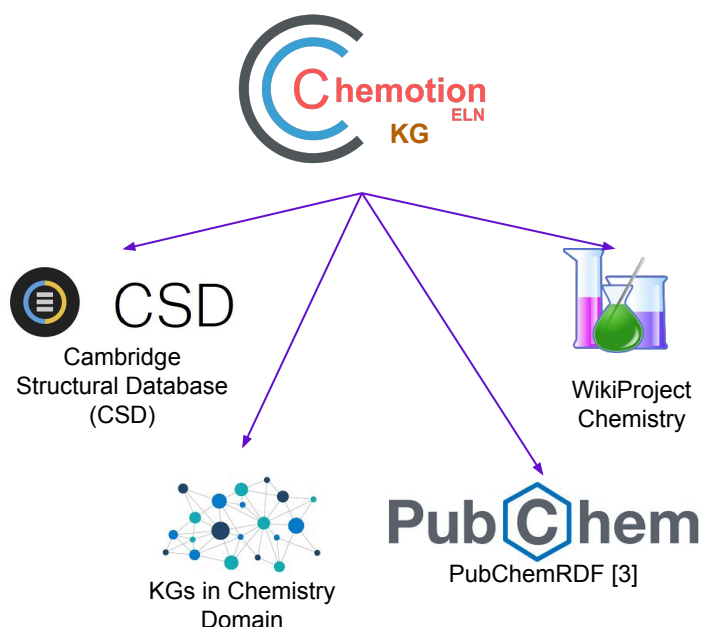


Knowledge-Guided Discovery in the Chemotion Knowledge Graph

Interdisciplinary topic co-supervised by KIT IBCS (Institute of Biological and Chemical Systems)



The *Chemotion* [1] Electronic Lab Notebook (ELN) serves as a digitalization instrument for work in chemistry and integrates state-of-the-art Research Data Management techniques into the routine work of academic researchers, thereby improving the data that is obtained as well as the quality of information that is published.

This thesis focuses on leveraging the *Chemotion Knowledge Graph (KG)* to extract and enrich chemical knowledge by integrating it with other chemical databases, knowledge graphs, and large language models (LLMs). It involves mapping entities across heterogeneous data sources, applying machine learning techniques for knowledge discovery, and evaluating how LLMs can complement structured scientific data. The work contributes to the broader goal of enhancing automated scientific reasoning and supporting FAIR (Findable, Accessible, Interoperable, Reusable) data principles in chemistry.

Tasks

- Conduct a literature review of chemical knowledge graphs, databases, and transformer-based LLMs.
- Design and implement entity alignment techniques to map Chemotion KG entities to external resources.
- Apply machine learning algorithms (e.g., clustering, classification) to identify patterns and derive insights.
- Explore integration of LLMs for KG completion, reasoning, and chemical question answering.

Sources

[1] Chemotion: <https://chemotion.net/>

[2] Chemotion KG: <https://dtrare.ise.fiz-karlsruhe.de/chemotion-kg/>

[3] PubChemRDF: <https://pubchem.ncbi.nlm.nih.gov/docs/rdf-federated-query>

Contact person

Ebrahim Norouzi¹

ebrahim.norouzi@fiz-karlsruhe.de

Co-supervisors

Anna Jacyszyn¹

anna.jacyszyn@fiz-karlsruhe.de

Jörg Waitelonis¹

joerg.waitelonis@fiz-karlsruhe.de

Nicole Jung²

nicole.jung@kit.edu

¹FIZ Karlsruhe, KIT AIFB

²KIT IBCS

