

Thesis & Project Topics

Markus Krötzsch
Knowledge-Based Systems

TU Dresden, 2026-10-22

Wikidata & Wikimedia



Wikidata

- Wikipedia's knowledge graph
- Free, community-built database
- Large graph (>1.7B statements on >120M entities)
- Large, active community
(several 100K logged-in human editors)
- Many applications

Wikimedia Deutschland e. V.

- German chapter of the Wikimedia Foundation
- Operates Wikidata
- Based in Berlin





- **Graph Rule Engine**
Executes Datalog (and extensions thereof) programs on Knowledge Graphs
- **Built for the Semantic Web**
Reads and writes RDF, queries SPARQL endpoints
- **Developed locally**
Free and open source, see <https://github.com/knowsys/nemo/>
- **Try it!**
<https://tools.iccl.inf.tu-dresden.de/nemo/>

Applying & Supporting Wikidata

How can the data in Wikidata be used?

(W1) **Wikidata reuse & community**

[Project/BSc](#): how is data from Wikidata currently reused by the community?

(W2) **Understanding the Wikidata ontology**

[BSc/MSc](#): which usage patterns exist within the Wikidata ontology?

(W3) **Schema extraction**

[Project/MSc](#): extract ShEx/SHACL schemas from sets of Wikidata items.

(W4) **Games using terms from Wikidata**

[Lab/Project](#): play, e.g., Codenames using terms from Wikidata.

User Platform

Work towards a collaborative platform for using Nemo.

(P1) **Visual output formats**

[Lab/Project](#): visual output formats (graphs, maps, etc.) for Nemo-web.

(P2) **Reactive Visualisation**

[Project/MSc](#): a reactive visualisation library in the style of **clinguin**, where UI is defined using Datalog facts.

(P3) **MediaWiki content model for rules**

[Project/BSc/MSc](#): extend MediaWiki to support editing and storing Nemo programs.

(P4) **Nemo Hibernation Format**

[BSc/MSc](#): design and implement a native hibernation format for Nemo.

(P5) **Investigate Wikifunctions**

[Project/BSc/MSc](#): How does **Wikifunctions** handle caching of intermediate results?

(P6) **Metadata from information extraction**

[Project/BSc/MSc](#): Which metadata can be derived during rule-based information extraction?

Extending Nemo

Implement functionality to extend the Nemo ecosystem.

(N1) **Reasoning Dashboard**

Lab: a Dashboard that displays what is currently happening during reasoning.

(N2) **BibTeX import/export**

Lab: allow Nemo to read and write BibTeX bibliographies.

(N3) **Active rule reasoning**

BSc/MSc: support user interaction during reasoning.

(N4) **Equality Reasoning**

Project/BSc/MSc: implement Equality Generating Dependencies.

(N5) **Graph data values**

BSc/MSc: a built-in datatype for representing graph values.

(N6) **HDT import/export**

Lab: allow Nemo to read and write Header-Dictionary Triples.

(N7) **Statistical analysis on term usage**

Project/BSc/MSc: Guide dictionary optimisation by studying term usage over time.

Reasoning with Datalog

Investigate extensions of Datalog, Nemo's rule language.

(D1) **Semantic diffing**

BSc/MSc: which changes to Datalog programs do not change the resulting facts?

(D2) **Comparing reasoning outputs**

Lab/Project/BSc/MSc: do reasoning outputs in different formats represent the same data?

(D3) **Generalising Safety**

BSc/MSc: how to propagate safety through built-ins and infinite IDBs?

(D4) **Modules & incremental computation**

Project/BSc/MSc: how to track data flow & push bindings across modules?

(D5) **Modules & stratification**

BSc/MSc: how to identify & repair non-stratified usage of modules?

(D6) **Certified stratified Datalog**

BSc/MSc: extend the certified Datalog implementation with stratified negation & stratified aggregates.

Applying Nemo

Apply Nemo to solve problems and investigate which functionality is needed.

(A1) **Bibliography validation**

Lab: validate & complete bibliographies by aggregating knowledge from, e.g., DBLP.

(A2) **Processing Data from Mass Spectrometry**

Lab/Project: identify molecules from mass spectrometry data.

(A3) **Benchmarking Rule Reasoners**

Project/BSc/MSc: develop a benchmark suite for Nemo and rule reasoners.

(A4) ...

Can I use AI?

- Yes
- Planned usage agreed with supervisor
- Detailed declaration required, including chat logs (supplementary material)
- **Your** work will be graded

PhD positions & PhD Fast Track