

zung der Daten wird erst in den vielfältigen und flexiblen Abfragemöglichkeiten deutlich und ist demnach nicht loszulösen von SPARQL.

SPARQL (SPARQL Protocol and RDF Query Language) ist eine 2008 vom W3C veröffentlichte, graphenbasierte Abfragesprache für RDF (Resource Description Framework). RDF ist ein Datenmodell, mit dem sich Ressourcen im World Wide Web darstellen lassen. Es ist der zentrale Standard des W3C, der semantische Daten in der charakteristischen Tripel-Struktur bestehend aus ‚Subjekt – Prädikat – Objekt‘ repräsentiert. Ausgehend von einem einzelnen solchen Tripel wird die Struktur eines Knowledge Graphen im Workshop entfaltet und die „Übersetzung“ von Forschungsfragen in natürlicher Sprache in die SPARQL-Syntax erläutert.

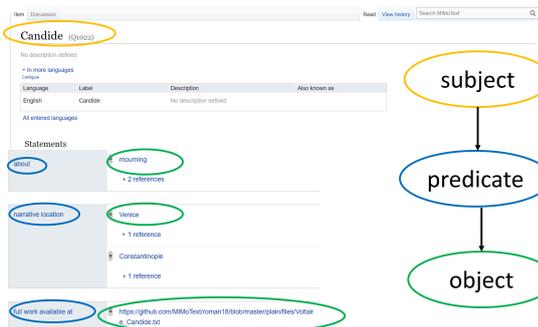


Abb. 2: Tripel-Struktur bestehend aus ‚Subjekt – Prädikat – Objekt‘, hier ein Beispiel zu einem literarischen Werk (Candide) aus der MiMoTextBase: <http://data.mimotext.uni-trier.de/wiki/Item:Q1022>.

Die Abfragesprache SPARQL setzt sich aus mehreren Bausteinen zusammen: *pattern matching* (Filtern des Datenbestands), *solution modifier* (Bearbeitung der Zwischenergebnisse) & *output* (Ausgabe als Tabelle oder Graph; Arenas et al. 2010). SPARQL ermöglicht es User:innen, durch Rekombination von Datensätzen neue Muster in den Daten zu erkennen und hypothesengeleitete Abfragen zu formulieren. Die Stärken dieser Abfragesprache und der Strukturierung von Wissen als RDF Triplestore sollen im Workshop in Anwendungsbeispielen gezeigt werden.

SPARQL-Abfragen werden häufig innerhalb eines einzelnen Knowledge Graphen gestellt. Es besteht jedoch auch die Möglichkeit, über mehrere Knowledge Graphen hinweg Abfragen zu stellen, sogenannte *federated queries* (Prud'hommeaux / Buil-Aranda 2013). Hier kommt das volle Potential von LOD zum Vorschein, denn so lässt sich Erkenntnisgewinn aus der Kombination mehrerer Graphen ziehen, ohne durch Replikationen von Datensätzen unnötige Redundanzen zu erzeugen. Im Workshop werden *federated queries* mit Wikidata erlernt und es wird gezeigt, inwieweit Nutzen aus diesen gezogen werden kann.

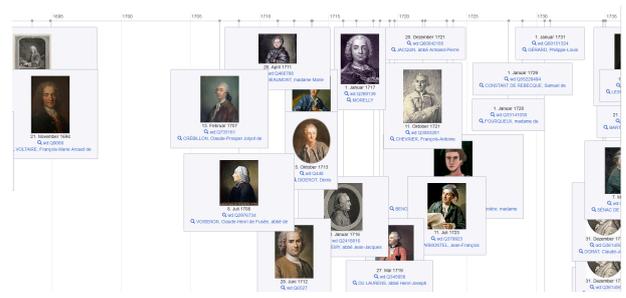


Abb. 3: *Federated queries* in SPARQL erlauben es, eigene Daten und externe Datenbestände (z. B. Wikidata) gleichzeitig abzufragen. Beispiel: <https://tinyurl.com/2ke42c3f>.

Workshop-Konzept

Der Workshop vermittelt Grundlagenwissen und Möglichkeiten, die das LOD-Paradigma bietet. Der im Projekt erstellte multilinguale Wissensgraph MiMoTextBase zur Domäne der französischen Literatur des 18. Jahrhunderts soll dabei als Anschauungsbeispiel dienen.

Lernziele

Der Workshop möchte praktisches Wissen vermitteln: Wie schreibt man SPARQL-Queries? Welchen Mehrwert kann ein Knowledge Graph für literaturgeschichtliche Fragen im Besonderen und die Geisteswissenschaften im Allgemeinen bieten?

In dem halbtägigen Workshop wird der Wissensgraph MiMoTextBase (Hinzmann et al. 2022a) des Projekts „Mining and Modeling Text“ vorgestellt und es werden Anwendungsszenarien formuliert und visualisiert. Dazu werden Grundlagen der Abfragesprache SPARQL erlernt und eigene Queries formuliert. Wir arbeiten beispielhaft auf dem projektinternen Knowledge Graphen sowie auf Wikidata und werden über *federated queries* die Verbindung mehrerer Wissensgraphen demonstrieren.

Konkrete Lernziele sind: Erwerb von Grundlagenwissen zu Semantic Web und RDF, LOD, Wikidata Graph; vertiefte Kenntnisse zu SPARQL und die praktische Fähigkeit, eigene SPARQL-Queries zu formulieren; Kennenlernen der Software Wikibase und Exploration der Visualisierungsmöglichkeiten des SPARQL-Endpoints.

Zielpublikum und Anforderungen

Der Workshop wendet sich an digitale Geisteswissenschaftler:innen mit Interesse an LOD und SPARQL. Spezielle Vorkenntnisse sind nicht notwendig. Teilnehmende benötigen einen Laptop.

Struktur / Ablauf

Der Workshop setzt sich aus aufeinander aufbauenden Sessions zusammen, die jeweils Input-Phasen und Übungsphasen verbinden. Es wird vorab eine ausführliche

che Tutorial-Seite (inklusive Verlinkung auf weitere hilfreiche Ressourcen zum SPARQL-Lernen) zur Verfügung gestellt, die den Teilnehmenden (und allen weiteren Interessierten) in der Vorbereitung sowie zur Vertiefung nützlich sein kann (Hinzmann et al. 2022b).

Im Zentrum des Workshops stehen drei Blöcke mit jeweils unterschiedlichem Schwerpunkt, in denen das Formulieren von SPARQL-Queries geübt wird (vgl. für Details den Ablauf im Appendix). Auch Teilnehmende ohne Vorkenntnisse werden schrittweise an zunehmend komplexere Queries herangeführt. Der Schwierigkeitsgrad wächst innerhalb der einzelnen Blöcke, wobei der Fokus auf dem eigenständigen Formulieren sowie Anpassen von Beispiel-Queries und dem Klären aller dabei auftretenden Fragen liegen wird.

1. Im ersten Teil liegt der Fokus auf Abfragen zu literarischen Werken. Im Hinblick auf SPARQL geht es hier zunächst um die zentralen Grundlagen wie das Schreiben einfacher *triple patterns* und Möglichkeiten der Kombination mehrerer *triple patterns* zu zunehmend komplexeren Queries. Der Mehrwert, der sich aus solchen Kombinationsmöglichkeiten ergibt, wird mit dem durch die MiMoTextBase gegebenen Fokus auf den französischen Aufklärungsroman besonders deutlich.

2. Im zweiten Teil widmen wir uns Wikidata als größtem öffentlichen Wissensgraphen, der sich zugleich als ‚Hub‘ begreifen lässt (Neubert 2017), und fokussieren Autor:innen als Entitäten. Autor:innen sind in allen geisteswissenschaftlichen Disziplinen relevant und ein wichtiges Scharnier zwischen verschiedenen Wissensgraphen. Bezogen auf die SPARQL-Syntax gehen wir einen Schritt weiter und integrieren Funktionen wie OPTIONAL und FILTER, um das Spektrum der Abfragemöglichkeiten zu erweitern. Ein Einstieg wird hier mit Queries zu Literatur:innen der MiMoText-Domäne gemacht. Im nächsten Schritt können die Teilnehmenden die Daten von Autor:innen in ihrer jeweiligen Domäne in Wikidata explorieren.

3. Der dritte Teil verknüpft die beiden vorigen Teile auf mehreren Ebenen. Der Schwerpunkt liegt auf *federated queries*, wobei wir uns auf Abfragen, die sich über die MiMoTextBase und Wikidata erstrecken, konzentrieren werden. Die genauere Betrachtung von Autor:innen des 2. Teils wird hier fortgesetzt und vertieft. In diesem abschließenden Teil wird der Mehrwert von Standards und geteilten Datenmodellen (Ontologien bzw. *entity schemata*) sowie die Verknüpfung von Ressourcen besonders deutlich.² Alle Autor:innen der MiMoTextBase, für die es auch Wikidata-Items gibt, können mit diesen über die Property *exact match* verknüpft werden, wodurch zusätzliche Informationen bereitstehen und diverse Abfragemöglichkeiten eröffnet werden.³ Die Wikibase-Infrastruktur bietet außerdem vielfältige Explorationsmöglichkeiten, die beispielhaft eingeführt werden (*marker cluster* für Geo-Daten, Timelines für Geburtsdatum u. ä.).

Es soll in der abschließenden Diskussion auch Raum sein, einen kritischen Blick auf Entwicklungen im Bereich des Semantic Web zu werfen, beispielsweise die Frage, welche Monopolisierungskräfte und Marktkräfte Einfluss nehmen (van Hooland / Verborgh 2014, 247–48; Singhal 2012). Zum Abschluss werden die wichtigsten Anwendungsmöglichkeiten und Fragen zusammengetragen und weiterführende Ressourcen (DuCharme 2013; van Hooland / Verborgh 2014; Lincoln 2015; Blaney 2017)

sowie bei Interesse Möglichkeiten der Kooperation thematisiert.

Appendix

Ablauf (4 Stunden)

10 Min.	Begrüßung (Vorstellung evtl. über Mentimeter)
20 Min.	Einleitung: Input zu Semantic Web, RDF, LOD für die Literaturgeschichte am Beispiel des Projekts Mining and Modeling Text, Wikidata & Wikibase Ecosystem, Mehrsprachigkeit des Graphen.
	SPARQL Teil 1 (MiMoTextBase)
20 Min.	(a) Input zu SPARQL-Grundlagen (interaktive Phase), SPARQL-Syntax, Möglichkeiten der Datenvisualisierung in Wikibase, Debugging & Help.
35 Min.	(b) Praxis-Teil: Anpassen vorhandener und Formulieren einfacher, eigener SPARQL-Queries auf der MiMoTextBase (Breakout Session bzw. Gruppenarbeit).
15 Min.	Pause
	SPARQL Teil 2 (Wikidata)
20 Min.	(a) Input: Erweiterte Elemente der SPARQL-Syntax wie OPTIONAL und FILTER; Datenmodell für Autor:innen auf Wikidata.
35 Min.	(b) Praxis: Formulieren etwas komplexerer Queries auf Wikidata.
15 Min.	Pause
	SPARQL Teil 3 (<i>Federated queries</i>)
20 Min.	(a) Input: Fortgeschrittene SPARQL-Queries: <i>Federated queries</i> , <i>prefixes</i> definieren, <i>marker cluster</i> etc.
35 Min.	(b) Praxis: <i>Federated queries</i> etc. anwenden.
15 Min.	Abschließende Diskussion, Empfehlung weiterführender Ressourcen zur Vertiefung des Gelernten.

Organisatorisches

Maximale Zahl der Teilnehmenden: 25. Wir benötigen einen Raum mit WLAN und Beamer und bieten gern ein Hybrid-Szenario an.

Fördernachweis

„Mining and Modeling Text“ (Universität Trier, Trier Center for Digital Humanities) wird von der Forschungsinitiative des Landes Rheinland-Pfalz 2019-2023 gefördert.

Beitragende

Der Workshop wird von Mitarbeiter:innen des LOD-Projekts „Mining and Modeling Text“ durchgeführt. Das interdisziplinäre Projekt verfügt über einen eigenen SPARQL-Endpoint und wurde in Wikibase implementiert.

Maria Hinzmann; hinzmannm@uni-trier.de; Trier Center for Digital Humanities, Universität Trier | Historisches Seminar: Digital Humanities, Bergische Universität Wuppertal; Forschungsinteressen: Datenmodellierung, LOD, Textanalyseverfahren.

Anne Klee; klee@uni-trier.de; Trier Center for Digital Humanities; Forschungsinteressen: Digitale Textverarbeitung; Digitale Lexikographie.

Johanna Konstanciak; konstanciak@uni-trier.de; Trier Center for Digital Humanities; Forschungsinteressen: Digitale Textverarbeitung; XML/Web-Technologien.

Julia Röttgermann; roettger@uni-trier.de; Trier Center for Digital Humanities; Forschungsinteressen: LOD, Textmining-Verfahren wie Topic Modeling, NER und Sentiment Analysis.

Christof Schöch; schoech@uni-trier.de; Trier Center for Digital Humanities; Forschungsinteressen: Computational Literary Studies.

Moritz Steffes; steffesm@uni-trier.de; Trier Center for Digital Humanities; Forschungsinteressen: Softwaresysteme, Semantic Web Technologien, Forschungsinfrastrukturen.

Fußnoten

1. Einige aktuelle Beispiele für Forschungsprojekte, die Wikibase verwenden: Enslaved (Zhou et al. 2020), Rhizome Artbase (Rhizome 2021), FactGrid (Simons 2022; Brunner 2022).
2. Diese Verknüpfung entspricht dem 5. Stern im Linked Open Data-Modell von Berners-Lee (2006).
3. Vgl. dazu den entsprechenden Query auf der MiMo-TextBase: .

Bibliographie

- Alves, Daniel, Hrsg.** 2022. „IJHAC: A Journal of Digital Humanities. Special Issue: Linked Open Data in the Arts and Humanities“ 16 (1). <https://www.eupublishing.com/doi/epdf/10.3366/ijhac.2022.0271>.
- Arenas, Marcelo, Claudio Gutierrez und Jorge Pérez.** 2010. „On the Semantics of SPARQL“. In *Semantic Web Information Management: A Model-Based Perspective*, herausgegeben von Roberto de Virgilio, Fausto Giunchiglia, und Letizia Tanca, 281–307. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-04329-1_13.
- Berners-Lee, Tim.** 2006. „Linked Data – Design Issues“. 27. Juli 2006. <https://www.w3.org/DesignIssues/Linked-Data.html>.
- Blaney, Jonathan.** 2017. „Introduction to the Principles of Linked Open Data“. *Programming Historian*. <https://programminghistorian.org/en/lessons/intro-to-linked-data>.
- Brunner, Katharina.** 2022. „FactGrid wants to become part of the Wikidata federation ecosystem“. 30. Mai 2022. <https://blog.factgrid.de/archives/2922>.
- Dörpinghaus, Jens.** 2022. „Wissensgraphen: Interdisziplinäre Perspektiven für Linked Data in den Geistes- und Sozialwissenschaften.“ *Zeitschrift für digitale Geisteswissenschaften* 07. https://doi.org/10.17175/2022_011.
- DuCharme, Bob.** 2013. Learning SPARQL. Sebastopol, UNITED STATES: O'Reilly Media.
- Hinzmann, Maria, Anne Klee, Johanna Konstanciak, Julia Röttgermann, Christof Schöch und Moritz Steffes. 2022a. „MiMoTextBase“, Trier Center for Digital Humanities, <https://data.mimotext.uni-trier.de, 11/2022>.
- Hinzmann, Maria, Anne Klee, Johanna Konstanciak, Julia Röttgermann, Christof Schöch und Moritz Steffes.** 2022b. „MiMoTextBase Tutorial“. Juli 2022. https://mimotext.github.io/MiMoTextBase_Tutorial/.
- Hogan, Aidan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard de Melo, Claudio Gutierrez, Sabrina Kirrane, et al.** 2021. „Knowledge Graphs“. *Synthesis Lectures on Data, Semantics, and Knowledge* 12 (2): 1–257. <https://doi.org/10.2200/S01125ED1V01Y202109DSK022>.
- Hooland, Seth van und Ruben Verborgh.** 2014. *Linked Data for Libraries, Archives and Museums: How to clean, link and publish your metadata*. London: Facet Publishing.
- Ikonić, Nevena, Milica, Ranka Stanković und Biljana Rujević.** 2021. „Serbian ELTeC Sub-Collection in Wikidata“. *Infotheca* 21 (2): 60–86. <https://doi.org/10.18485/infotheca.2021.21.2.4>.
- Lincoln, Matthew.** 2015. „Using SPARQL to access Linked Open Data“. Herausgegeben von Fred Gibbs. *The Programming Historian*, Nr. 4 (November). <https://doi.org/10.46430/phen0047>.
- Neubert, Joachim.** 2017. „Wikidata as a Linking Hub for Knowledge Organization Systems? Integrating an Authority Mapping into Wikidata and Learning Lessons for KOS Mappings“. In *Proceedings of the 17th European NKOS workshop*. <http://ceur-ws.org/Vol-1937/paper2.pdf>.
- Ohmukai, Ikki und Taizo Yamada, Hrsg.** 2022. *Digital Humanities 2022. Conference Abstracts. Responding to Asian Diversity*. Tokyo: ADHO. <https://dh2022.dhii.asia/dh2022bookofabsts.pdf>.
- Prud'hommeaux, Eric und Carlos Buil-Aranda.** 2013. „SPARQL 1.1 Federated Query“. W3C Recommendation. 21. März 2013. <https://www.w3.org/TR/sparql11-federated-query/>.
- Rhizome.** 2021. „The ArtBase Relaunches: Welcome to Linked Open Data. Rhizome“. <http://rhizome.org/editorial/2021/apr/26/the-artbase-relaunches-welcome-to-linked-open-data>.
- Sack, Harald und Mehwish Alam.** 2020. „Knowledge Graphs“. Potsdam. <https://open.hpi.de/courses/knowledgegraphs2020>.
- Schöch, Christof, Maria Hinzmann, Röttgermann Julia, Anne Klee und Katharina Dietz.** 2022. „Smart Modelling for Literary History“. *IJHAC: International Journal of Humanities and Arts Computing [Special issue on Linked Open Data]* 16 (1): 78–93. <https://doi.org/10.3366/ijhac.2022.0278>.
- Simons, Olaf.** 2022. „FactGrid“, Forschungszentrum Gotha der Universität Erfurt, database.factgrid.de, 11/2022. <http://doi.org/10.17616/R31NJMQR>.
- Singhal, Amit.** 2012. „Introducing the Knowledge Graph: Things, Not Strings“. *Google (blog)*. 16. Mai 2012. <https://blog.google/products/search/introducing-knowledge-graph-things-not/>.
- Thornton, Katherine, Kenneth Seals-Nutt, Mari- anne Van Renmoortel, Julie M. Birkholz und Peterjan De Potter.** 2021. „Linking Women Editors of Periodicals to the Wikidata Knowledge Graph“. *Semantic Web journal Special Issue Cultural Heritage* 2021. <http://www.semantic-web-journal.net/content/linking-women-editors-periodicals-wikidata-knowledge-graph>.
- Zhao, Fudie.** 2022. „How to Critically Utilise Wikidata - A Systematic Review of Wikidata in DH Projects“. In *Digital Humanities 2022 - Conference Abstracts*, herausgegeben von Ikki Ohmukai und Taizo Yamada, 608–10. The University of Tokyo, Japan: DH2022 Local Organizing Committee. <https://dh2022.dhii.asia/dh2022bookofabsts.pdf>.
- Zhou, Lu, Cogan Shimizu, Pascal Hitzler, Alicia M. Sheill, Seila Gonzalez Estrecha, Catherine Foley, Duncan Tarr und Dean Rehberger.** 2020. „The Enslaved Da-

taset: A Real-world Complex Ontology Alignment Benchmark using Wikibase“. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 3197–3204. CIKM '20. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3340531.3412768>.