

Wikidata Fortgeschrittenen-Kurs

Vertiefung von Wikidata

Bern, 10. November 2021

Dieser Kurs kann dank der Unterstützung folgender Organisationen durchgeführt werden:





Inhalte

Teil 1: Antworten zu Fragen aus dem Einsteiger-Kurs Teil 2: Konzept Linked Open Data

Teil 3: SPARQL-Queries

Teil 4: Automatisierter Datenimport



Kursblock 1: Antworten zu Fragen aus Wikidata Einsteiger-Kurs

Was ist ein Lexem?



Eine vollständige Anleitung für Anfängerinnen und Anfänger gibt es unter Gemeinschaftsportal.

Erfahre mehr über Wikidata

Wikispecies Wikibooks

Wikimania

Wikinews

Wikipedia

- Was ist Wikidata? Lies die Wikidata-Einführung.
- Erkunde Wikidata, indem du dir das Vorzeigeobjekt über den britischen Schriftsteller Douglas Adamsig anschaust.

Aktuelle Highlights

Dean Stockwell (Q3452)

Was ist ein Lexem

 Lexikalisches Element einer Sprache, das in etwa der Menge der flektierten Formen eines einzelnen Wortes entspricht

• Lexeme sind Entitäten im Sinne des Wikibase-Datenmodells

● z.B. Lexem "<u>singen</u>" (=L48427)

• Aber <u>Sänger</u> (L48427) und <u>Sängerin</u> (L73023) sind eigene Lexeme!

Mehr Informationen:

https://www.wikidata.org/wiki/Wikidata:Lexicographical_data/Documentation/de#Lexem

https://de.wikipedia.org/wiki/Lexem

Was bezweckt die "Service-Line" in einer SPARQL-Query?

Wikidata Query Service		Wikidata Query Se	ervice		
 i SELECT ?child WHERE i (there is a child in the state is a child		 SELECT ?child ?c WHERE S S (4 ?child father 7. > ?child wit:P22 6 SERVICE wikiba 7) 9 9 8 	: Bach : Wd:Q13 se:labe	el 39. 1 { bd:servicsParam wikibase:language "en". }	
	0				0
© - 0 20 Epsebases in 11 ms ↔ Code 🛓 Heruntertaden -	Ø Link≁		_	20 Ergebnisse in 197 ms 🚸 Code 🛃 Herunterladen 🗸 🔗 L	ink -
child	¢	child	¢	childLabel	4
Q wd:Q57225		Q wd:Q57225		Johann Christoph Friedrich Bach	T
Q wd:Q76428		Q wd:Q76428		Carl Philipp Emanuel Bach	
Q wd:Q106641		Q wd:Q106641		Johann Christian Bach	
Q wd:Q107277		Q wd:Q107277		Wilhelm Friedemann Bach	
Q wd:Q470198		Q wd:Q470198		Johann Gottfried Bernhard Bach	

Das Ergebnis ist enttäuschend: Du siehst nur die Identifikatoren der Ergebnisse. Du kannst auf diese Identifikatoren klicken und dann die dazugehörige Wikidata-Seite (mit verständlichen Beschriftungen) öffnen. Aber gibt es nicht eine elegantere Art, die Ergebnisse anzuzeigen?

Um die rhetorische Frage zu beantworten: Ja! Mit der zauberhaften Zeile:

```
SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE]". }
```

die du irgendwo innerhalb der WHERE -Klausel hinzufügst: Für jede Variable ?foo gibt es nun einen value ?fooLabel, der das Etikett für ?foo beinhaltet. Wenn Du also diese Variable der SELECT -Klausel hinzufügst, wird auch das Etikett angezeigt.

```
SELECT ?kind ?kindLabel
WHERE
{
    # ?kind vater Bach
    ?kind wdt:P22 wd:Q1339.
    SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE]". }
}
```

Versuche es!

Versuche die Abfrage auszuführen. Du solltest nun nicht nur die Nummern sehen, sondern auch die Namen der Kinder.

kind	kindLabel
wd:Q57225岱	Johann Christoph Friedrich Bach
wd:Q76428🗗	Carl Philipp Emanuel Bach



Refresher



Kursblock 2: Konzept Linked Open Data

Linked Open Data

Bezeichnet im World Wide Web frei verfügbare Daten

• Sind per Uniform Resource Identifier (URI) identifiziert

• Können per HTTP abgerufen werden

Linked Open Data ist Teil des Semantic Web

URI, URL, URN

Beispiel

- t3n.de URI
- https://t3n.de URL (http zeigt euch wo die Ressource ist)
- ftp://t3n.de URL (ftp zeigt euch wo die Ressource ist)
- urn:isbn:3827370191 URN (eindeutige Indentifikation des Buches "Moderne Betriebssysteme" von Andrew S. Tanenbaum)

Grundkonzept

Mit Linked Open Data soll ein Netz aus Daten entstehen

• Geht im Wesentlichen auf Tim Berners Lee

• Er prägte vier Regeln für Linked Data:

- Verwende zur Bezeichnung von Objekten URIs
- Verwende HTTP-URIs, so dass sich die Bezeichnungen nachschlagen lassen
- Stelle zweckdienliche Informationen bereit, wenn jemand eine URI nachschlägt (mittels der Standards RDF und SPARQL)
- Zu diesen Informationen gehören insbesondere Links auf andere URIs, über die weitere Objekte entdeckt werden könne



Mehr infos unter: https://5stardata.info/de/

eCH Standard

Dieses <u>Dokument</u> bietet der Schweizer Linked-Data-Community eine gemeinsame Vision zum Stand der Veröffentlichung von verknüpften offenen Daten im öffentlichen Sektor und im Kulturerbe in der Schweiz und gibt Personen, die neu in der Community sind, einen ersten Überblick über frühere und laufende Aktivitäten im Bereich der Datenveröffentlichung, der Datennutzung und des Wissensaustauschs.

Das Dokument enthält eine kurze Einführung in die Thematik der Linked (Open) Data, gibt einen detaillierten Überblick darüber, worum es bei der Veröffentlichung von Linked Data geht, gibt einen Überblick über den derzeitigen Stand der Veröffentlichung von Linked Data durch öffentliche und kulturhistorische Organisationen in der Schweiz und stellt eine Reihe von beispielhaften Anwendungsfällen vor, die als Test- und Studienfälle dienen, um aktuelle Herausforderungen zu bewältigen und den Nutzen von Linked (Open) Data in der Praxis zu demonstrieren.



Kursblock 3: SPARQL queries

Zur Erinnerung

Akronym - steht für SPARQL Protocol And RDF Query Language

• Ist die Standardsprache, um als RDF Triple verfasste Graph-Daten abzufragen.

Orientieren an den Beispielen im Query Service

• Offizielle Website: <u>https://www.w3.org/TR/sparql11-overview/</u>

Autovervollständigung

• Vorschläge mit Strg+Leertaste (oder Alt+Eingabe oder Strg+Alt+Leertaste)

• Wähle den Code mit der Pfeiltaste aus und bestätige mit Enter

• SERVICE wikibase:label { bd:serviceParam wikibase:language "en". } = SERV

• Zeitsparend für unsere Abfragen

ORDER BY und LIMIT

• Oft ältestes oder jüngstes Ergebnis am interessantesten

ORDER BY sortiert Ergebnisse nach irgendwas

LIMIT begrenzt Suche

• Hilft uns bei der Suche

Karte

• Erstellen von Karten für bessere Visualisierung

Werden durch Daten von OpenStreetMap unterstützt

Weiterführende Links zu SPARQL

https://docs.data.world/tutorials/sparql/Your_First_Sparql_Query.html

https://www.wikidata.org/wiki/Wikidata:SPARQL_tutorial/de

https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service/queries/examples/de

https://joinup.ec.europa.eu/sites/default/files/document/2015-05/d2.1.2_training_module_1.3_introduction_to_rdf_sparql_v1.00_de.pdf

https://www.mediawiki.org/wiki/Wikidata_Query_Service/User_Manual/de



Kursblock 4: Automatisierter Datenimport

Datenspende für Wikidata



Gründe um Daten auf Wikidata hochzuladen

- Mehr Leute haben Zugriff auf die Daten
- Das freie Wissen wird verbessert und erweitert
- Führt zu einem höhren Web Traffic
- Macht die Daten nützlicher

Erstellen eir Zu importie Starten eines Starten eines Mikidata Dafinieren o Wikidata Daten Daten Hinzufügen Hinzufügen Prüfen des	ines Wikidata Accounts erende Daten auswählen erende Daten auswählen at der Daten in ein att n der zu importierenden r Importmethode für die er Importmethode für die ar laten mit Wikidata r der Daten zu Wikidata Datenimports
--	--

In 10 Schritten zum Datenimport

Schritt o: Erstellen eines Wikidata Accounts

• Um einen Datenimport auf Wikidata starten zu können, benötigt man einen Wikidata Account.

• Wie man einen solchen erstellt, wurde im Einsteiger-Kurs diskutiert.

Schritt 1: Zu importierende Daten auswählen

In Wikidata importierte Daten müssen...

...zuverlässig sein.

...öffentlich verfügbar sein, vorzugsweise online.

Schritt 2: Starten eines Datenimports

Wikidata:Dataset Imports

tart a data import	
1. Create an account or log in	
1. Create an account by clicking 'Create an account' in the top right hand corner of the page, please enter an email address to allow Wikidata users to notify you about discussions on the dataset).	
2. Log in by clicking 'Log in' in the top right hand corner of the page.	
2. Click the Start a data import button	Start a data import
3. Replace NAME OF THE DATASET with the name of the dataset in the Subject field	
 Fill in the preloaded fields Name of dataset, Source, Link and Description. 	
5. Click 'Publish changes'.	
6. Follow the instructions given here to launch the data import.	
7. Delete the import demand from the list below once the data are imported in Wikidata and add it to the imported data sets page.	

Process [Bearbeiten]

Importing datasets through Wikidata:Dataset Imports has a number of benefits over importing the dataset without documenting the process. The process has 5 stages:

Identify datasets: There are many external services that can help identify datasets on a subject including Google Dataset Search 🕏, Wolfram Data Repository 🕏 and Humanilarian Data Exchange 🕄

In 10 Schritten zum Datenimport

Auf <u>Wikidata:Dataset Imports</u> wirst du durch den Prozess des Datenimports geführt. Bitte befolge die Anweisungen, um einen neuen Import zu starten.

Stelle sicher, dass der Datensatz so klar und gründlich wie möglich beschrieben wird. Bitte tue dies auch, wenn du die Daten schlussendlich selbst importierst. Mit einem guten Beschrieb stellst du sicher, dass auch andere aus der Community mithelfen können und verstehen, was du für zukünftige Aktualisierungen der Daten getan hast.

Dataset name

The name of the dataset

e.g. Museums of São Paulo

Source

The name of the organisation who produces the dataset

e.g. Sao Paulo State Secretariat of Culture

Link

A link to the dataset. If the dataset has an information page, please link with the full URL to that.

e.g. http://www.wikidata.org

Dataset description

A description of the dataset, e.g what information is held within it

e.g. List of 485 museums of São Paulo, the state of Brazil with the biggest number of museums

Dataset format

×

The format of the dataset, if it is a list on a website, a speadsheet a pdf, etc.

Dataset topic

The topic that most closely fits this dataset, if there is no clear match please chose General or Unknown

w.

Updates

Is the dataset likely to be updated? Either the same dataset is revised and changed, or new versions are created e.g statistics on a subject published each year

		+

Additional information

Is there any additional information that can help describe the dataset?



Schritt 3: Importieren der Daten in ein Tabellenblatt

Es wird dringend empfohlen, dass du die Daten in eine Online-Tabelle (z. B. auf einer GAFAM-Website wie Google Sheets oder auf einer Community-Website wie <u>Framacalc</u>) importierst und dort weiterbearbeitest. Dies ermöglicht es anderen aus der Community, die Probleme zu verstehen und beim Import eines Datensatzes mitzuarbeiten. Es ist sehr hilfreich, ID-Nummern anzugeben, wenn diese für den Upload-Prozess vorhanden sind.

Wenn die Daten nur als Tabelle in einer PDF-Datei vorliegen, kann <u>Tabula</u> sehr hilfreich sein um die Daten aus der PDF-Datei in eine Online-Tabelle zu importieren.

GAFAM = Big Tech Companies (Google, Apple, Facebook, Amazon, Microsoft)

Schritt 4: Definieren der Struktur der Daten in Wikidata

Dieser Schritt ist oft der schwierigste. Es gibt aber viele sachkundige Personen in der Wikidata-Community, die mit dir zusammenarbeiten können, um diesen Schritt zu bewältigen. Auf der Seite <u>Wikidata:Partnerships and data imports</u> kannst du deine Fragen stellen (vorzugsweise auf Englisch).

Part A: Schaue dir das Wikidata-Glossar an, um die in den folgenden Schritten verwendeten Begriffe zu verstehen.

Part B: Suche nach Beispielen von potenziell ähnlichen Daten in Wikidata, damit du verstehts, welche Struktur verwendet wird.

• <u>Vorzeigeobjekte</u> sind Datenobjekte, die vollständig zeigen, wozu Wikidata in der Lage ist und detailliert die Daten zeigen, die in Wikidata gespeichert und möglicherweise in allen Wikimedia-Projekten und außerhalb von Wikimedia genutzt werden können

• Verwende die Suchfunktion, um nach Objekten zu suchen, die möglicherweise ähnliche Informationen enthalten, die so gespeichert sind, dass sie für diesen Datensatz kopiert werden könnten.

© Überprüfe die Liste der Eigenschaften, um Eigenschaften für deine Daten zu finden oder neue Eigenschaften vorzuschlagen.

Schritt 5: Formatieren der zu importierenden Daten

Part A: Dupliziere die Datei des Originaldatensatzes in deinem Tabellenkalkulationsprogramm und benenne die Kopie in "structured for Wikidata" um.

Part B: Verwende nun die von Wikidata vorgegebene <u>Struktur</u> für dein Tabellenblatt. So erleichterst du es den Personen aus der Community, dich beim Importieren deiner Daten in Wikidata zu unterstützen. Eine entsprechende Vorlage findest du <u>hier</u>.

Schritt 6: Auswahl der Importmethode für die Daten

Grundsätzlich gibt es zwei Optionen:

- 1. Deine Daten werden von anderen Personen aus der Community importiert.
- 2. Du importierst die Daten selbst.

Datenimport durch Personen aus der Community

Schritt A: Beantrage den Import von Daten in Wikidata auf der Seite <u>Wikidata bot request</u>. Um einen Antrag zu stellen, klicke auf die Schaltfläche "Einen neuen Antrag stellen" und verlinke dann deine Anfrage auf der Data Hub-Seite.

Schritt B: Überprüfe den Abschnitt "manual work needed" in der Tabelle, sobald der Import begonnen hat, um zu sehen, welche Arbeiten manuell durchgeführt werden müssen.

Schritt C: Gehe zu Schritt 9, sobald die Daten durch die Community importiert wurden.

Selbständiger Datenimport

Bevor du Daten über eine beliebige Artikelliste importieren kannst, musst du die entsprechenden Wikidata-ID-Nummern für jeden Artikel in der Liste (im Wesentlichen jede Zeile deines Arbeitsblatts) kennen.

Wie das geht wird dir im Schritt 7 erklärt.

Schritt 7: Abgleich der Daten mit Wikidata (1/2)

- a) Suche nach den Nummern der Wikidata-Datenobjekte für jede "Zeile" in den Daten (z. B. die Nummer für jedes Gemälde, wenn der Datensatz eine Sammlung von Gemälden ist). Für diesen Schritt hast du folgende Möglichkeiten:
 - i) <u>Mix'n'match</u> Verwende dieses Tool, wenn ein grosser Teil der Daten manuell von Community-Mitgliedern abgeglichen werden muss.
 - ii) <u>OpenRefine</u> Verwende dieses Tool für den präzisen Abgleich unter Verwendung mehrerer Spalten deines Datensatzes. Entweder durch Fuzzy-Matching, externe Identifikatoren oder Seitenlinks.
 - iii) Anderer halbautomatischer Abgleich Wenn die Daten einige eindeutige IDs, URLs oder andere Daten mit Wikidata gemeinsam haben, die es dir ermöglichen, eine Liste von Q-Nummern zu generieren (z. B. durch Ausführen einer Abfrage und Abgleich gemeinsamer Daten in Ihrer Import-Tabelle).

Schritt 7: Abgleich der Daten mit Wikidata (2/2)

b) Abgleich der als Werte verwendeten Elemente:

> Ermittlung der Q-Nummern für alle zusätzlichen Daten, die hinzugefügt werden müssen. Ein Gemälde mit "Herkunftsland" = "Italien" muss z. B. zu P495 = Q38 werden (Hinweis: Die Eigenschaften sollten bereits in Schritt 5 gefunden worden sein).

Die Methode, die du dafür verwendest, hängt vom jeweiligen Fall ab. Die folgenden Werkzeuge stehen zur Verfügung:

- i) OpenRefine kann dafür wieder verwendet werden; ii)
 - Wikipedia und Wikidata Tools bietet eine ähnliche Funktionalität für Google Sheets.

 \rightarrow Bei allen Methoden kann es sein, dass du am Ende noch etwas Handarbeit leisten musst.

Schritt 8: Hinzufügen der Daten zu Wikidata

Nach dem Abgleich der Daten stehen mehrere Optionen für den Import zur Verfügung:

- <u>OpenRefine</u> zur Umwandlung von Tabellendaten in Wikidata-Edits
- <u>PetScan</u> Für den Fall, dass Sie dieselbe(n) Aussage(n) zu einer Liste von Artikeln hinzufügen müssen.
- <u>QuickStatements</u> Für Aussagen, Bezeichnungen, Beschreibungen usw., die sich von Artikel zu Artikel unterscheiden.
- Bot-Import Wenn der Auftrag zu komplex ist, um QuickStatements oder PetScan zu verwenden. Kannst du <u>hier</u> einen Bot-Import anfordern.

Schritt 9: Prüfen des Datenimports

Sobald die Daten in Wikidata importiert wurden, generierst du eine Abfrage (=Query), um sicherzustellen, dass deine Daten korrekt importiert wurden (Wikidata:Request a query).

Entweder kannst du die Abfrage direkt selbst im Wikidata Query Service generieren,

oder du lässt dir <u>hier</u> von der Community helfen.

Schritt 10: Zusammenfassen des Datenimports

Part A: Füge Informationen in den Abschnitt "Date import complete and notes" der Tabelle auf der Seite für den Datensatzimport ein und füge ein Datum hinzu, an dem der Datenimport in Wikidata abgeschlossen wurde, sowie Anmerkungen zu Problemen oder anderen Dingen, die man wissen sollte, um die Daten verbessern oder pflegen zu können (z. B. wann eine aktualisierte Version des Datensatzes veröffentlicht wird).

Part B: Verschieben den Datensatz aus dem Abschnitt "<u>In Progress datasets</u>" in den Abschnitt "Abgeschlossene Datensätze".

Beispiele von Zusammenfassungen

https://www.wikidata.org/wiki/Wikidata:Dataset Imports/Higher Education Institutions of Brazil (2011)

https://www.wikidata.org/wiki/Wikidata:Dataset Imports/Interlinking dataset of the video game databases Mobygames, GameFAQs and MediaArt_DB#Getting_help

https://www.wikidata.org/wiki/Wikidata:Dataset Imports/Fertility rate for countries and geographical regions

Anleitung von Beat Estermann (2017)

https://www.wikidata.org/wiki/Wikidata:WikiProject Cultural heritage/Guidelines/Ingesting Datasets into Wikidata#Tools:

Herzlichen Dank an unsere Projektpartner!



