



Wikidata Einsteiger-Kurs

Einführung in Wikidata

Bern, 20. Oktober 2021

Dieser Kurs kann dank der Unterstützung folgender Organisationen durchgeführt werden:



MY-D.ORG

Inhalte

Teil 1: Einführung

Teil 2: Konzepte und Syntax

Teil 3: Einfaches Erfassen von Daten

Teil 4: Einfache SPARQL-Abfragen

Kursblock 1: Einführung

Wikimedia

"Die Wikimedia Foundation ist die gemeinnützige Organisation, welche die Wikipedia und andere **Projekte für Freies Wissen** beherbergt. Wikimedia will es für alle einfacher machen, ihr Wissen mit anderen zu teilen." https://wikimediafoundation.org/de/about

- weltweite Bewegung zur Förderung Freien Wissens
- gemeinnützige Stiftung mit Sitz in San Francisco
- mehr als 450 Mitarbeitende
- ehrenamtliches Engagement steht im Vordergrund

Wikimedia Projekte

Wikipedia

Wikimedia Commons

Wikivoyage

Wikibooks

Wikiversity

• • •

Wikidata



Wikidata - Was, Wie, Warum?

Wikidata

- ist eine frei bearbeitbare und frei zugängliche Wissensdatenbank in der eine Vielzahl von Informationen in Form von Triples gespeichert wird.
- ist nicht dasselbe wie Wikipedia, steht aber mit Wikipedia in Verbindung
- hat zum Ziel
 - o die Verlinkung zwischen den verschiedenen Wikipedia Sprachversionen zu zentralisieren
 - o einen zentralen Speicherort für <u>Infoboxdaten</u> aller Wikipedien bereitzustellen

Triples

Was ist ein **Triple**? Es ist die kleinste faktische Aussage, die gemacht werden kann, um Informationen über etwas festzuhalten. Zum Beispiel:

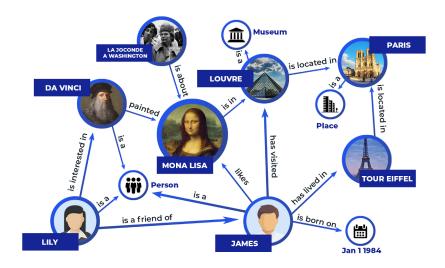
"Mona Lisa — wurde gemalt von — Leonardo Da Vinci"

→ Subjekt — Prädikat — Objekt

Eine Datenbank, die aus Triples besteht wird "knowledge base" (=Wissensdatenbank) genannt.

Knowledge Base grafisch darstellen

Triples können graphisch dargestellt werden. Man nennt dies einen **Knowledge Graph**.



 $\frac{\text{https://yashuseth.blog/2019/10/08/introduction-question-answering-knowledge-graphs-}{\text{kgqa/}}$

Infoboxen

Infoboxen sind eine Möglichkeit zur einheitlichen und einfachen Verwendung von Tabellen, die bei einigen Artikeln am Anfang stehen und grundlegende Daten enthalten. Infoboxen werden mit Formatvorlagen jeweils speziell für eine bestimmte Reihe von Artikeln (z. B. alle Berge, Bands ...) gestaltet.

Die Infoboxen sollen ein anschauliches Hilfsmittel zum Fliesstext sein und diesen nicht ersetzen, sondern lediglich ergänzen.



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

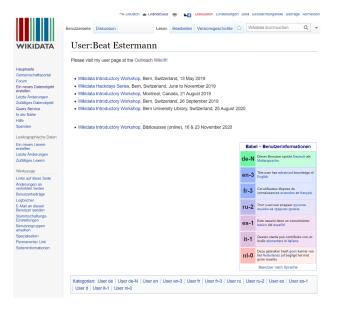
Wikidata Account

Wer in Wikidata editieren will, benötigt einen Account.

Jeder User soll seinen eigenen Account eröffnen.

Es empfiehlt sich eine E-Mail-Adresse mit dem Account zu verknüpfen.

Innerhalb des Wikidata Accounts können die User Informationen über ihre Person teilen.



Wikidata Account - Beat Estermann: https://www.wikidata.org/wiki/User:Beat_Estermann

Task 1: Wikidata Account eröffnen (10 Min)

- ▶ Username: Wähle ein Pseudonym, das nicht deinem Künstlernamen oder Firmennamen ähnelt.
- ► Lasse dich vom User Account von Beat Estermann inspirieren.

Beachte: Du kannst für Wikidata, Wikipedia und alle anderen Wikimedia Dienste denselben Account benutzen. Falls du also bereits einen Account eines anderen Wiki-Dienstes haben solltest, kannst du dich bei Wikidata mit dem bereits existierenden Benutzerkonto einloggen.

Kursblock 2: Konzept und Syntax

Zwei Hauptziele von Wikidata

Die Zentralisierung der Sprachversionen-Interwikilinks zwischen den Wikipedia-Projekten

und

das Bereitstellen einer allgemeinen Wissensquelle für die gesamte Welt.

Vier Schlüsselkonzepte

- (Daten-)Objekt oder Entität (Englisch: item or entity)
- **Eigenschaft** (Englisch: property)
- Wert (Englisch: value)
- Aussage (Englisch: statement)

Diese und weitere Begriffe werden im Glossar genauer erklärt.

(Daten-)Objekt

- Repräsentiert einen Gegenstand, Konzept oder Ereignis in der wirklichen Welt (konkret oder abstrakt).
- Ihm wird in Wikidata ein Identifikator (äquivalent zum Namen) und Informationen zugeordnet.
- Jedes Objekt hat eine zugehörige Wikiseite und eine einzigartige ID beginnend mit "Q" (z.B. Mona Lisa = Q12418)
- Der Identifikator macht das (Daten-)Objekt maschinenlesbar.
- Der Teil mit den eigentlichen Daten ist die Liste der Aussagen über das Objekt.
- Ein Objekt kann als Subjekt eines Triples in Linked Data betrachtet werden.

Eigenschaft

- Ist eine Beschreibung für einen Datenwert einer Aussage.
- Kann als Datenkategorie gedacht werden, zum Beispiel "Farbe" für den Datenwert "blau".
- Eigenschaften bilden in Wikidata zusammen mit Werten eine Aussage.
- Eigenschaften haben eigene Seiten auf Wikidata (beginnend mit dem Buchstaben P)
- Dementsprechend resultiert eine verlinkte Datenstruktur → Linked Data.



https://www.wikidata.org/wiki/O12418

Wert

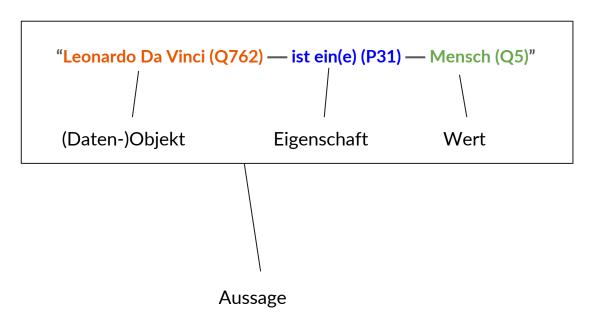
- Werte sind sind die Informationen, die in einer Behauptung (=claim) gespeichert werden.
- Werte können in verschiedenen Datentypen angegeben werden (z.B. GEO Koordinaten, URL, Zeit, etc.).
- Werte können aber auch andere (Daten-)Objekte sein.

Aussage

- Eine Aussage entspricht einem Triple.
- Ein Triple besteht aus
 - · einem (Daten-)Objekt,
 - einer Eigenschaft und
 - einem Wert

<u>Bemerkung</u>: Wikidata macht keine Annahmen über die Richtigkeit von Aussagen, sondern sammelt und meldet sie lediglich mit einem Verweis an eine Quelle. Der Begriff "Aussage" wird häufig synonym mit "Behauptung" verwendet, technisch wird eine Behauptung jedoch nur zu einer Aussage, wenn wenigstens eine Fundstelle angegeben wurde.

Übersicht Schlüsselkonzepte



Leonardo da Vinci (Q762)

italienischer Maler, Bildhauer, Architekt, Anatom, Ingenieur und Naturphilosoph (1452-1519) Leonardo di ser Piero | Lionardo di ser Piero | Leonardo di ser Piero da Vinci

In weiteren Sprachen

Aussagen





https://www.wikidata.org/wiki/Q762

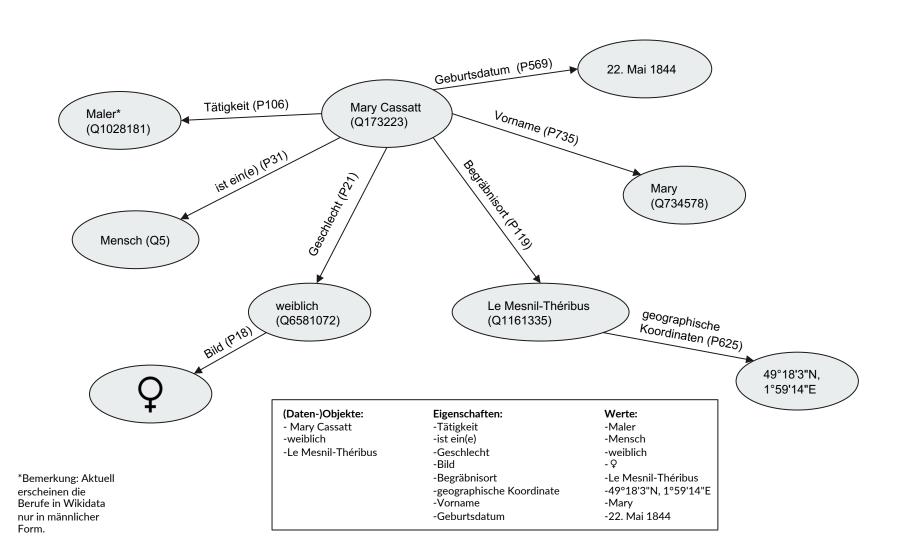
Triples/Aussagen: Verschiedene Wörter, gleiche Konzepte

Triples und Aussagen sind dieselbe Methoden um Wissen darzustellen. Jedoch ist ihre Terminologie

verschieden. Rahmen der Wissensrepräsentation	Referenzobjekt	Link	Wert
Resource Description Framework (RDF)	Subjekt	Prädikat	Objekt
Wikidata Aussage	(Daten-)Objekt	Eigenschaft	Wert
Alltagssprache	Ding	Beziehung	Ding

Task 2 - Vertiefen der Syntax (10 + 5 Min)

- Denke dir ein (Daten-)Objekt aus deiner Branche.
- ▷ Überlege dir 5 Aussagen zu diesem (Daten-)Objekt und erstelle basierend darauf deinen eigenen Knowledge-Graphen.
- ➤ Ergänze nun deinen Knowledge-Graphen mit zwei bis drei Aussagen in denen ein Wert des ursprünglich gedachten (Daten-)Objekt selbst zum (Daten-)Objekt wird.
- Orientiere dich am Beispiel auf der n\u00e4chsten Folie







Kursblock 3: Einfaches Erfassen von Daten in Wikidata

Überlegungen vor einem Wikidata-Objekt-Eintrag

Relevanz:

- D.h. mindestens 1 Link zu einem verwandten Wikimedia-Projekt,
- eine klar identifizierbare konzeptionelle oder materielle Einheit, die durch seriöse und öffentlich zugängliche Referenzen beschrieben werden kann oder ein Artikel der ein strukturelles Bedürfnis erfüllt.
- Weitere Informationen findest du in den Wikidata-Relevanzkriterien.

• Einmaligkeit:

- Wikidata-Objekte sollen einmalig sein.
- Jedes Objekt sollte ein eindeutig identifizierbares Konzept oder eine Sache repräsentieren (oder eine Ausprägung davon).
- Konzept (Planet), Ausprägung des Konzepts (Erde)

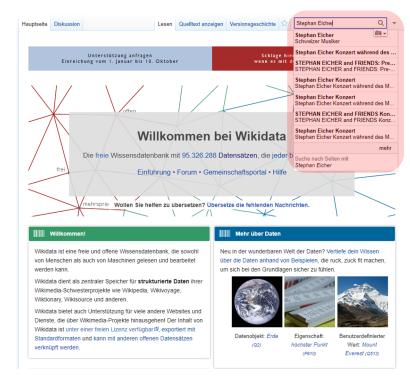
Verknüpft:

- Objekte können und sollen (sofern sinnvoll) miteinander verknüpft werden.
- Auf den Wikidata-Seiten der Objekte können Verknüpfungen hinzugefügt werden.

Erstellen eines (Daten-)Objekts

Überprüfen, ob das (Daten-)Objekt bereits vorhanden ist:

- Suche nach dem (Daten-)Objekt, das du erstellen möchtest.
- Wenn es mehrere Artikel unter demselben Namen gibt, musst du evtl. die Liste durchsuchen um sicherstellen zu können, dass das Objekt noch nicht existiert.
- Beginne nur mit dem Erstellen des neuen (Daten-)
 Objekts, wenn es noch nicht in Wikidata existiert.



Erstellen eines (Daten-)Objekts

Hauptseite

Forum

erstellen

Gemeinschaftsportal

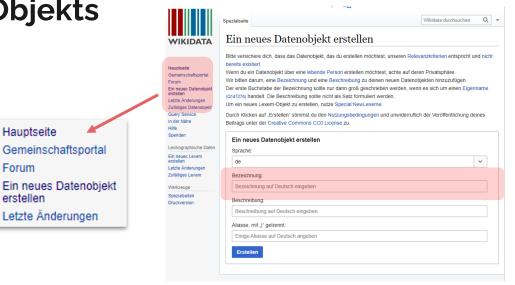
Letzte Änderungen

Im linken Menii kannst du stets den Button "ein neues Datenobjekt erstellen" auswählen.

Unter dem Listeneintrag "Bezeichnung" kannst du beispielsweise den Namen der/des Künstler(in/s) hinzufügen.

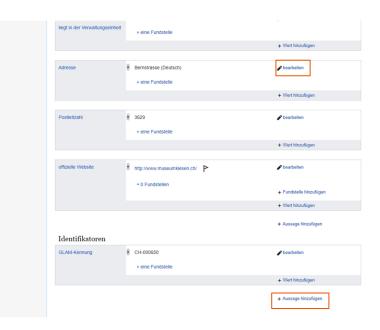
Danach klickst du auf die Schaltfläche "Erstellen".

Keine Sorge wegen den anderen Einträgen, diese kannst du zu einem späteren Zeitpunkt vornehmen.



Bearbeiten/Aktualisieren eines (Daten-)Objekts





Task 3 - Erstellen/Bearbeiten eines (Daten-)Objekts (15 Min)

- ▶ Falls es noch nicht vorhanden ist, erstellst du dieses.
- > Falls es bereits vorhanden ist, bearbeitest du dieses, indem du eine Aussage hinzufügst.





Kursblock 4: Einfache SPARQL-Abfragen

Was ist SPARQL?

- Akronym steht für für SPARQL Protocol And RDF Query Language
- Ist die Standardsprache, um als RDF Triple verfasste Graph-Daten abzufragen.
- Offizielle Website: https://www.w3.org/TR/sparql11-overview/
- Eine umfangreichere Einführung gibt es im Folgekurs am 10. November 2021

Wozu verwendet man SPARQL in Wikidata?

- Mit SPARQL können wir Abfragen (Queries) für die Wikidata-Wissensdatenbank formulieren.
- Anders formuliert: Mit SPARQL können wir Fragen an Wikidata Stellen.
- Mit korrekt formulierten Queries können wir schnell Informationen über die Datenbank gewinnen.
- Z.B. können wir nach allen (Daten-)Objekten suchen die der Instanz "Ziege" zugeordnet sind.
- Es können aber auch weit komplexere Abfragen konstruiert werden, so dass auch scheinbar komplizierte Fragen rasch beantwortet werden können.

Was ist eine Abfrage (Query)?

- Eine Abfrage ist eine spezielle Form, eine Frage so zu formulieren, dass ein Computersystem sie verstehen und beantworten kann.
- Damit dies gelingt müssen die Abfragen nach einem genauen Schema aufgebaut sein.
- Dieses Schema nennt man auch Syntax der Abfragesprache.

Vier verschiedene Arten von Abfragen

- SELECT-Abfrage
- CONSTRUCT-Abfrage
- ASK-Abfrage
- DESCRIBE-Abfrage

SELECT-Abfrage

- Die SELECT-Abfrage wird verwendet, um Rohwerte aus einem SPARQL-Endpunkt (in unserem Fall die Wikidata-Datenbank) zu extrahieren.
- Die Ergebnisse der SELECT-Abfrage werden in einem Tabellenformat zurückgegeben.
- In unserem Einsteiger Kurs widmen wir uns nur der SELECT-Abfrage.

Der WHERE-Block

- Im WHERE-Block werden Suchkriterien definiert, die die Abfrage einschränken.
- Wie bei einem Filter in einem Tabellenkalkulationsprogramm können wir definieren, welche Teilmenge wir uns anzeigen lassen möchten.
- Alle Abfrageformen (ausser DESCRIBE-Abfrage) benötigen zwingend einen WHERE-Block.

Schema der SELECT-Abfrage

```
SELECT ?frucht
WHERE
{
    ?frucht farbe gelb.
    ?frucht geschmack sauer.
}
```

- Ein Triple ist wie ein Satz aufgebaut: Es hat ein "Subjekt", ein "Prädikat" und ein "Objekt". Und er hört mit einem Punkt auf.
- Wir suchen nach dem Subjekt, dessen Prädikate die im WHERE-Block definierten Werte enthalten.
- Das Beispiel liest sich wie folgt:

"Gebe mir alle (Daten-)Objekte vom Typ "Frucht" aus, deren Farbe gelb und deren Geschmack sauer ist."

Korrekte Syntax der SELECT-Abfrage

- Damit die SELECT-Abfrage maschinenlesbar wird, dürfen wir nur Begrifflichkeiten verwenden, die in der Syntax definiert sind.
- Die im vorherigen Beispiel definierten Einschränkungen "farbe gelb" und "geschmack sauer" sind in dieser Form nicht maschinenlesbar.
- Stattdessen müsste "farbe" & "geschmack" eine P-Nummer und "gelb" & "sauer" eine Q-Nummer sein.
- Ausserdem müssen wir die Wikidata-Präfixe wd: (für Elemente) und wdt: (für Eigenschaften) verwenden.

Wikidata Query Service

- Mit dem Wikidata Query Service kannst du dir einen Überblick über die Daten in Wikidata verschaffen.
- Hier kannst du deine Abfragen programmieren und ausführen.

Eine korrekte SELECT-Abfrage

```
#Hausziegen
SELECT ?Hausziege ?HausziegeLabel

WHERE
{
    #ist ein(e) Hausziege
    ?Hausziege wdt:P31 wd:Q2934.
    SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
}
```

Hausziege	HausziegeLabel
Q wd:Q151345	Geißbock Hennes
Q wd:Q3569037	William Windsor
Q wd:Q23003932	Old Whiskers
Q wd:Q24287064	Taffy
Q wd:Q41239734	Lance Corporal Shenkin III
Q wd:Q41240892	Lance Corporal Shenkin II
Q wd:Q41241416	Lance Corporal Shenkin I

Beispiele adaptieren

Der Wikidata Query Service stellt Beispiele zur Verfügung, die

einfach adaptiert werden können:



Wikidata Query Builder

- Der Wikidata Query Builder bietet eine visuelle Schnittstelle zur Erstellung einfacher Wikidata-Abfragen.
- Er ist ideal für Benutzer mit wenig oder keiner SPARQL-Erfahrung.
- Der Query Builder bietet nicht die volle Funktionalität von SPARQL.
- Du kannst jedoch deine im Query Builder erstellte Abfrage jederzeit im Query Service öffnen, wo du deine Abfrage weiter bearbeiten und erweitern kannst.
- Mier geht es zum Wikidata Query Builder.

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

Task 4: Erstellen von einfachen SELECT-Abfragen (10min)

Abschluss

Danksagung

Einige Folien dieser Präsentation wurden direkt von <u>Linked Digital Future</u> übernommen und in die Deutsche Sprache übersetzt.

Besten Dank an Frédéric Julien und sein Team von Linked Digital Future für das Zurverfügungstellen der Folien!

<u>Hier</u> findest du die Aufzeichnungen und Folien vergangener Wikidata-Online-Kurse von Linked Digital Future und kannst dich ausserdem für künftige Kurse einschreiben.

Wikidata Fortgeschrittenen-Kurs

Am Mittwoch 10. November (13:30-15:15) findet der Wikidata Fortgeschrittenen Kurs statt.

Folgende Themen behandeln wir:

- SPARQL Abfragen kreieren
- Einführung in Linked Open Data
- Batch upload

Wir würden uns freuen, dich wiederum begrüssen zu dürfen!