

Wikidata

Steiner, Christian; christian.steiner@uni-graz.at

Wikidata wurde 2012 von der *Wikimedia Foundation* ins Leben gerufen, um eine gemeinsame Wissensbasis für strukturierte Daten für die verschiedenen Sprachen in *Wikipedia*, *Wikimedia Commons* und anderen Stiftungsprojekten zu schaffen. *Wikidata* teilt ihren gemeinschaftsbasierten Ansatz mit *Wikipedia*, wo jede und jeder Daten erweitern und bearbeiten kann. Die Wissensdatenbank (*knowledge base*) basiert auf den Prinzipien von *Linked Open Data* und benutzt die Software Wikibase als Backend.

Wikidata trennt in seinem Datenmodell *statements*, *items*, *properties* und *lexemes*.

Statements sind ganz im Sinne von RDF in *Triples* abbildbar, obwohl sie formal in *Wikibase* *key-value*-Paare sind. Sie bestehen aus einem *property*, dem ein oder mehrere Werte zugeordnet sein können und das sich auf ein *item* bezieht (z. B.: *University of Graz* [Q622683] [Q303] *instance of* [P31] *university* [Q3918], übersetzt: "Die Universität Graz ist eine Universität.").

Items repräsentieren Konzepte – von konkreten Objekten oder Themenfeldern bis hin zu immateriellen Ideen. Jedes *item* wird durch eine eindeutige Nummer identifiziert, der der Buchstabe Q vorangestellt wird und die gemeinhin als 'QID' bezeichnet wird. Dies ermöglicht eine sprachunabhängige Identifizierung der Konzepte. Beispiele für *items* sind Lebensmittel [Q2095], Liebe [Q316], Apfel [Q89] oder Universum [Q1]. Die *labels*, also semantische Bezeichnungen der *items*, müssen nicht eindeutig sein. Es gibt zum Beispiel zwei Artikel mit dem Namen 'Elvis Presley': Elvis Presley [Q303] steht für den amerikanischen Sänger und Schauspieler und Elvis Presley [Q610926] steht für sein selbstbetitelt Album. Grundsätzlich besteht ein *item* aus einem *label*, einer *description*, optional mehreren Aliasnamen und einer gewissen Anzahl von *statements*.

Ein *property* beschreibt den Datenwert eines *statements* und kann als eine Kategorie von Daten verstanden werden oder als Kante im Graph. Beispiele für *properties* sind etwa *instance of* [P31], *subclass of* [P279] oder *color* [P462].

Lexemes sind in *Wikidata* *items* mit einer Datenstruktur speziell für lexikographische Daten.

Wikidata hat sich mittlerweile zu einer multidisziplinären, maschinenlesbaren, zentralisierten und vernetzten Wissensdrehscheibe entwickelt und wird für immer mehr Anwendungsfälle genutzt. Es bietet sich für Digitale Editionen an, das historische Wissen in und über das edierte Werk über *Wikidata*-QIDs mit dem *Semantic Web* zu verknüpfen.

Literatur:

- Farda-Sarbas, Mariam; Müller-Birn, Claudia: Wikidata from a Research Perspective - A Systematic Mapping Study of Wikidata. In: ArXiv abs/1908.11153: 2019.
- Lemus-Rojas, Mairelys; Pintscher, Lydia: Wikidata and Libraries: Facilitating Open Knowledge. In: Leveraging Wikipedia: Connecting Communities of Knowledge. Chicago, IL: 2018, S. 143–158.
- . In: Wikidata: A platform for data integration and dissemination for the life sciences and beyond Wikidata: 2015.
- Waagmeester, Andra; Lynn, Schriml; Su, Andrew: Wikidata as a linked-data hub for Biodiversity data. In: Biodiversity Information Science and Standards 3: 2019.
- Wikidata. URL: <http://www.wikidata.org>

Software:

Wikidata, BlazeGraph, Wikisource

Verweise:

Linked Open Data, RDF, Normdaten, Semantic Web, GND, GeoNames, VIAF, Getty

Projekte:

CoReMA - Cooking Recipes of the Middle Ages, MEDEA. Modelling semantically Enriched Digital Edition of Accounts, Mittelhochdeutsche Begriffsdatenbank (MHDBDB), FactGrid - a database for historians

Themen:

Annotation und Modellierung, Archivierung, Digitale Editionswissenschaft

Zitervorschlag:

Steiner, Christian. 2021. Wikidata. In: KONDE Weißbuch. Hrsg. v. Helmut W. Klug unter Mitarbeit von Selina Galka und Elisabeth Steiner im HRSM Projekt "Kompetenznetzwerk Digitale Edition". URL: <https://gams.uni-graz.at/o:konde.112>