Wikidata's 2018/19 Highlights

Wikidata had an eventful year. Two completely new types of data have been opened up with the support for Lexicographical Data on Wikidata and Structured Data on Commons, additional improvements and new tools ensure that Wikidata's data stays of high-quality and several new Wikibase instances contribute to a growing Wikibase Ecosystem.

What are your highlights?

Lexicographical data:

Wikidata started out collecting data about concepts in Items. But there is more out there! There are words in thousands of languages. These words and their different forms and meanings can now be described in Lexemes on Wikidata. In the first year 55.000 Lexemes in 350 languages have been created. They are the start of an incredibly useful resource for machine translation, automated text generation and more. Bringing data about concepts and words together in the same place opens up many new opportunities.

Structured Data on Commons:

Wikimedia Commons is a treasure trove of amazing multimedia content. Over the last year the technical foundation has been set for making the data about it machine-readable. This will help make the content available on Wikimedia Commons much more discoverable and easier to re-use. Wikidata's Items and Properties make this possible by providing the multilingual tagging vocabulary necessary for machines to understand the information about images, videos, etc.

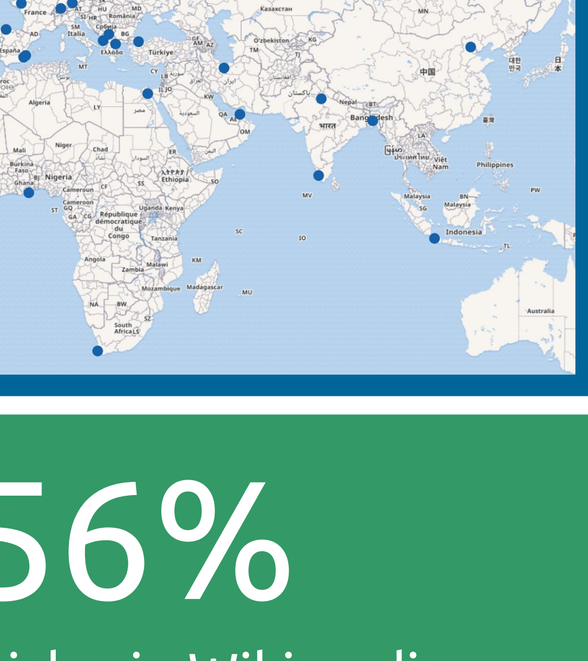
Data quality:

Several improvements have been made that help ensure Wikidata's data quality:

- Wikidata now supports Schemas. With this editors can define how certain areas of the data should look like using the ShEx standard. A list of Items can be tested against a Schema to find problematic Items. This enables wiki projects to have more control over the Items that matter to them in a systematic way.
- The constraint system has been extended. Now it is possible to define which entity types (i.e. Item, Property, Lexeme, Form, Sense, MediaInfo) a Property can be used on and ensure that two things happened within the same time-frame. Additionally a part of the constraint violations can be found in the Query Service, making it easy to create worklists of problematic Items.
- When making a new Statement that has a limited set of recommended values (e.g. driving side: left/right) these are now suggested, leading to fewer mistakes.
- For Statements with coordinates a map is shown to make it easier to spot potential errors in this type of data.

Wikibase Ecosystem:

Wikidata holds general-purpose data about the world. A lot of additional data should be available but not necessarily in Wikidata itself. Thanks to the possibility of running their own Wikibase instance more and more institutions and projects are opening up their data and connecting it to Wikidata.
Among them are LinguaLibre for recording sound files and FactGrid collecting data related to the humanities.
The Wikibase Registry at wikibase-registry.wmflabs.org gives an overview of them.
They are all interconnected in the Wikibase Ecosystem.



6th Birthday:

The Wikidata Community celebrated the 6th birthday of the project with more than 30 parties and meet-ups all over the world.

31.100
editors are editing on Wikidata each month

58 million
Items in Wikidata are describing the world

>56%
of articles in Wikimedia projects make use of data from Wikidata

Contact: Lydia Pintscher (lydia.pintscher@wikimedia.de) / @wikidata