Wikidata: Structured data repository that anyone can edit

Filip Maljković
Wikimedia Serbia, Belgrade, Serbia
filip@vikimedija.org

Abstract

Wikidata is an open collaborative multilingual knowledge base that deals with structured data. It is based on facts and references, it allows links to other databases and it is made for both humans and machines. It was created as a central repository of structured data for Wikimedia projects, especially Wikipedia, the well-known free online encyclopedia, so it is by extension also free: it is released under Creative Commons Zero license.

One of the most important purposes of Wikidata is answering questions that couldn't be answered before, such as "What are the biggest cities with a female mayor?". For querying the Wikidata database, users can use SPARQL (semantic query language for RDF – Resource Description Framework), and there are various tools for representing queried data in many visualization forms.

The other important usage of Wikidata is to have a centralized repository for data that could be used on Wikipedia, both in infoboxes that are later used by Google and other crawlers, but also in article text. That way, for instance, updating census information for countries could require changing data solely on Wikidata, without the need to update articles of all the cities in all the language versions on Wikipedia.

Since most users of Wikidata are machines, the purpose of humans is to create content and annotate it properly. For most people that might seem tedious, so several web-based games have been developed to help out with gamifying Wikidata, thus providing quality data without people having to go through the process of directly editing Wikidata pages.

Keywords: data; OER; open repository; collaboration.