Wikidata visualization by predefined queries related to one particular topic. A case study

Abstract:
We will map out a way in which information from Wikidata on a particular topic can be presented to a general public. This approach is that users do not need to have any prior technical knowledge for data retrieval and at the same time gain comprehensive insight into a specific topic. For data retrieval and at the same time gain comprehensive insight into a specific topic.

Wikidata has become one of the most important storages for free and open structured data. The database can be read and edited by both humans and machines. But if you are human it is not easy to retrieve aggregated data out of Wikidata. Because in general, the SPARQL query language is quite complex and difficult to understand. What is needed is a set of predefined SPARQL templates. The queries are hierarchically related to each other. The advantage of this approach is that users do not need to have any prior technical knowledge for data retrieval and at the same time gain comprehensive insight into a specific topic.

Our way:
The service called thesaurus is part of the online project "Archive guide to the German Colonial Past". The aim of the project is to provide an overview of archival documents related to the topic and which are held in different archives, museums and libraries. Furthermore, it should help to link these documents to information on places, actors, objects and events. The project has been funded with support from the German Foreign Office. A beta version was published in February 2019.

All objects of the data set are related to the main topic object "German colonial empire" (Q329618). The top level is made up of the individual colonies and the German Empire. The elements of this layer are specified manually. Below the entries for each individual colony there is a set of links. These links lead to queries about a certain class of objects related to this colony (e.g. persons, companies, missions, events, art works etc.). Further sub-queries are possible, e.g. for the staff of a particular Mission Society or the officials of a particular administrative unit. The results of these queries can also be sorted by time. As a result, users receive a chronology of events.

A link on the sidebar gives users access to the respective query in the Wikidata SPARQL query service. There they may modify the queries and export the results. If available, references are listed for each data object found. This provides an initial overview of the reliability of the data displayed.

Since the data is not cached on the project website but is obtained directly from Wikidata, users have the opportunity to supplement the data in Wikidata and complete the result lists further. Via a link on the sidebar it is possible to track all recent changes to the data corpus and thus detect vandalism.

The resulting data can be linked in several ways to the actual project database of archival descriptions. On the one hand, a full text search in the database is initiated via the link ‘Relevant Documents’. Labels and AlternateLabels in several languages are used for this purpose. On the other hand, objects with an existing statement "archived at" (P485) have an additional link. This link initiates a combination of free text search and initiates a combination of free text search and links to inventory descriptions in which the respective person or corporation acts as creator.

By using templates, we are currently able to offer around 8,000 different queries and show their results to the users.

The tool was written in Javascript/jQuery. The source code is available on the page.