Recommendation-Assisted Data Curation for Wikidata

Eva Zangerle[1], Claudia Müller-Birn[2]

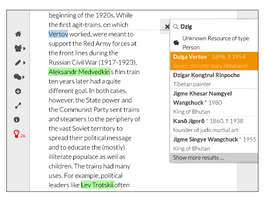
Motivation

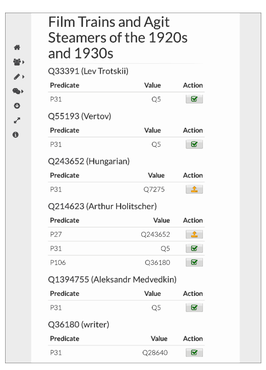
The quality and completeness of data contained in structured knowledge bases is naturally influenced by the users who enter and maintain the data in the form of items described by statements on the platform. Users who are new to the Wikidata environment and its underlying data model, but are, nonetheless, experts in their fields, are confronted with a steep learning curve when aiming to enter information on Wikidata (e.g. regarding the choice of suitable properties for creating statements).

We propose a recommendation-based annotation platform where users who currently work with or on a text are supported in finding suitable Wikidata entities for data extracted from the underlying text source to ultimately feed this structured information to the Wikidata platform. Such recommendations not only support users in annotating their data according to Wikidata's terminological knowledge, but also expand the number of references on the Wikidata platform that reveal the origin of existing statements.

Recommendation-Assisted Data Curation

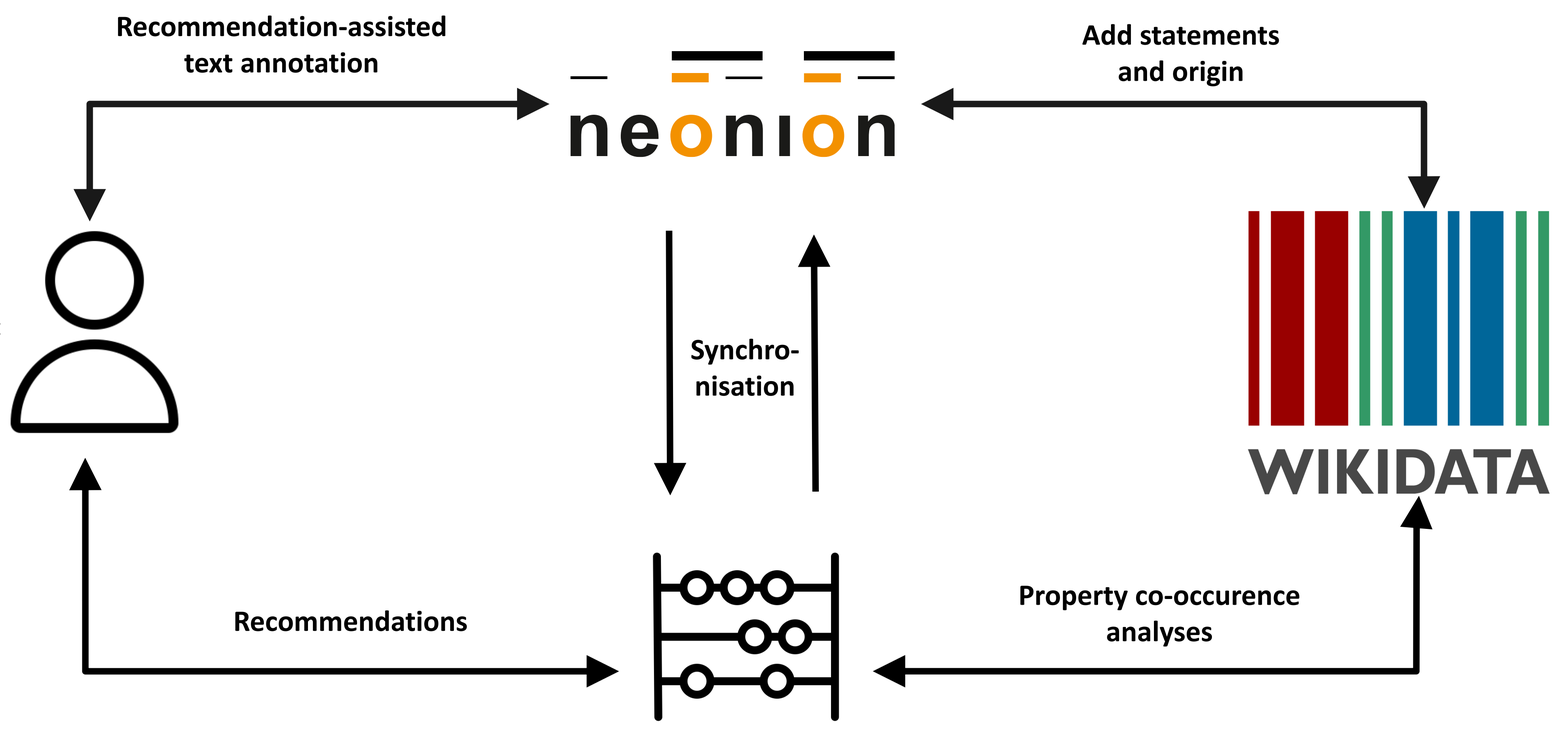
Our goal is to support user in annotating texts and feeding these annotations to Wikidata platform





Screenshots depicting the process shown in neonion's annotating interface.

neonion is a browser-based user interface to add semantic annotations in a triple format, where the text is the verifiable basis for statements that are added to Wikidata.



The snoopy system is a recommender system for properties on the Wikidata platform. It computes recommendations for suitable Wikidata properties based on the co-occurrence of properties on Wikidata items.

Contributions

- We present a prototype that allows users to annotate texts semantically, while being supported by recommendations that point the user to suitable properties and, ultimately, submit these structured annotations to Wikidata.
- We discuss how such an approach of providing information to enhance the verifiability of Wikidata's knowledge graph.
- We elaborate on future extensions to this prototype which aim to enhance the user experience and foster human-computer collaboration.

Future Work

- **Text Context:** named entity recognition of underlying texts for recommending parts that should be annotated, complemented with properties to be used for the recommendation.
- **Personalization*:*** extend recommender system to adapt recommendations to user's experience, preferences, etc.
- **Transparency and Feedback:** explain and justify recommendations to educate the user regarding Wikidata's data model.
- **Evaluation:** perform a user-centric evaluation of the proposed prototype.

[1] Department of Computer Science
Universität Innsbruck, Austria
https://dbis.uibk.ac.at
eva.zangerle@uibk.ac.at

[2] Institute of Computer Science
Freie Universität Berlin, Germany
https://hcc.mi.fu-berlin.de
clmb@inf.fu-berlin.de

Find our paper on https://doi.org/10.5281/zenodo.1194790