

# Call for Master thesis

## „Developing a text annotation tool for the knowledge graph Wikidata“

### What is the topic?

*Entity linking* (or *text annotation*) is the process of linking phrases in a text (e.g., “Trump”) to the corresponding entry in a knowledge graph (e.g., to [https://en.wikipedia.org/wiki/Donald\\_Trump](https://en.wikipedia.org/wiki/Donald_Trump) in the English Wikipedia). This helps for having a better understanding of what is written in the text (improving search) and, in contrary, to backup concepts (such as Donald\_Trump in Wikipedia) with references in a text.

In the past, several entity linking systems have been developed. However, they work mainly for Wikipedia and not for other knowledge graphs.

In this thesis, we focus on Wikidata [1,2] as knowledge graph. Wikidata is has emerged in recent years as a large knowledge graph containing more concepts than Wikipedia. Moreover, Wikidata is language-independent and stores useful meta-information, such as the source of facts. All this allows Wikidata to be a promising knowledge graph for the future.

In this thesis, the aim is to develop an entity linking system for Wikidata. Based on an already existing prototype implementation, the student needs to study the peculiarities of Wikidata and of its data model. Based on that, a new approach for entity linking will be developed and implemented, which performs better (using precision and recall, but also w.r.t. other qualitative aspects) than entity linking based on Wikipedia. The idea is to show that linking to Wikidata provides additional benefits. This will be demonstrated by an evaluation performed on benchmark data sets for Wikipedia and/or by a manual evaluation.

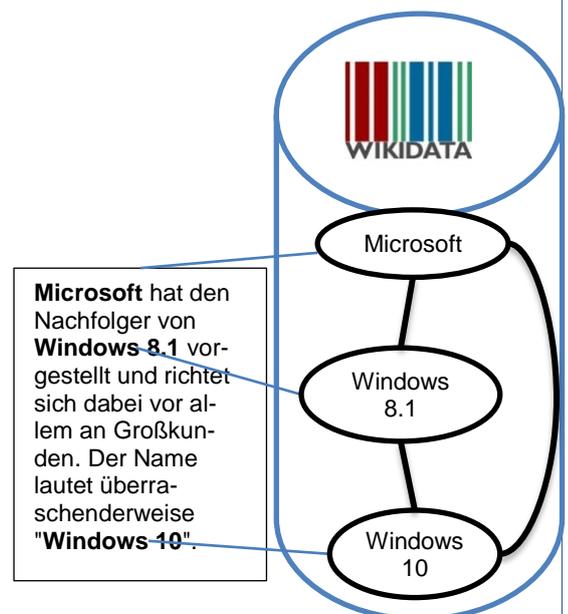
[1] <http://wikidata.org>

[2] <https://tools.wmflabs.org/reasonator/>

### Which prerequisites should you have?

- Interest in Semantic Web Technologies (knowledge graphs, maybe RDF)
- Knowledge in supervised machine learning can be beneficial.
- Good programming skills in Java.

Keywords: Implementation, entity linking, text annotation, knowledge graph, Wikidata, Natural Language Processing, Machine Learning



Contact person:  
**Michael Färber**  
[michael.farber@cs.uni-freiburg.de](mailto:michael.farber@cs.uni-freiburg.de)