

Nov. 14, 2014

Call for Bachelor/Master thesis “Wikification based on Wikidata”

What’s it all about?

Wikification is the process of adding Wikipedia links to phrases in a given text (text annotation). For that purpose some robust approaches have been developed recently. The Wikipedia Miner [1] is one example where the linking is learned by features which were extracted from Wikipedia and with a machine learning approach.

In this thesis, the technique of wikification shall be applied on other knowledge bases such as Wikidata [2][3]. This means that the features for linking phrases in a text to entities in the knowledge base must be taken from Wikidata and not Wikipedia. The linking is done to Wikidata and not Wikipedia. For comparison reasons it can be interesting to link also to Wikipedia articles, since most Wikidata entities also have a Wikipedia page.

For the thesis, a suitable set of features based on the Wikidata dataset is chosen. A Wikification system for Wikidata is implemented and evaluated. For that purpose, open source code such as of the Wikipedia Miner can be reused.

[1] <http://wikipedia-miner.cms.waikato.ac.nz>

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

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

What should you bring along?

- Interest in Semantic Technologies.
- Good programming skills in Java.

Contact person:
Michael Färber
michael.farber@kit.edu
Tel.: 0721/608 479 46