

07.04.2015

# Master-/Diplomarbeit „Automatische Textannotation mit Wikidata, einer Wissensbasis der Zukunft“ zu vergeben!

## Um was geht es?

Als *Wikification* wird das Vorgehen verstanden, bei dem Text mit Links zu Wikipedia versehen wird (sog. Textannotation) – etwa, um zu wissen, welche Dinge im Text vorkommen, um anschließend danach suchen zu können. Für Wikipedia wurden bereits erfolgreich einige Verfahren entwickelt. Hierzu zählt der Wikipedia Miner [1], bei dem die Verlinkung durch *features*, die aus Wikipedia extrahiert wurden, und einem überwachten *machine learning*-Verfahren erreicht wird.

In der Abschlussarbeit soll das Prinzip des Wikifizierens auf die Wissensbasis Wikidata [2][3] angewendet werden. Dies bedeutet zum einen, dass die *features* für das *Entity Linking* aus Wikidata – und nicht aus Wikipedia – extrahiert werden; zum anderen, dass nach Wikidata verlinkt wird.

Neben der Auswahl und der Extraktion der richtigen *features* basierend auf dem Wikidata-Datensatz besteht die Arbeit in der Implementierung und Evaluation eines *Wikidatafi-cation*-Systems. Hierbei kann auf dem Open-Source Code des Wikipedia Miner aufgebaut werden.

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

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

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

## Was sollten Sie mitbringen?

- Interesse an semantischen Technologien und *supervised machine learning*.
- Gute Programmierkenntnisse in Java.

Kontaktperson:  
**Michael Färber**  
michael.farber@kit.edu  
Tel.: 0721/608 479 46

