

## Call for Master thesis

# „No time for reading related work? A system for automatically enriching texts with citations“

### What is the topic?

Many authors face the challenge to add the optimal references to their scientific publications, such as conference papers. Due to the extremely increasing number of publications (about hundreds of thousand publications in the computer science domain only) and due to the rise of new research fields and new methods, it is more and more difficult and time consuming for authors to have an overview of the literature and to select the best references for their own texts.

This Master thesis focuses on the conceptualization and implementation of a tool for adding (“annotating”) references to a given input text (see figure on the right).

First of all, a large collection of publications is analyzed and the references with their citation contexts (e.g., the words standing next to the reference marker “[1]”) are extracted and represented in a suitable data model.

In the application phase, a new text should be enriched with citations. For that, the system checks which of the citation contexts stored in the model can be found in the text. For these found citation contexts, the corresponding references are ranked and the most important references are added to the text.

For the analysis and for building the model, necessary data (like corpora of publications) will be provided. An initial list of literature (for knowing related work) will also be given to the student.

### Which prerequisites should you have?

- Good programming skills, preferable in Java.
- Interest in the application of Text Mining/Natural Language Processing methods (e.g., named entity recognition, entity linking, word embeddings) and/or methods for disambiguating and ranking publications (using bibliometrics).
- Machine Learning skills a strong plus.

Keywords: Implementation, Natural Language Processing, Text Mining, Machine Learning, Information Retrieval, Bibliometrics, Recommender Systems.

... by researchers in the field of information retrieval (IR) ... In the popular tf-idf scheme, a basic vocabulary of words is chosen.



*Citation Recommendation Tool*



... by researchers in the field of information retrieval (IR) (**Baeza-Yates and Ribeiro-Neto, 1999**).... In the popular tf-idf scheme (**Salton and McGill, 1983**), a basic vocabulary of words is chosen.

Contact:

**Michael Färber**

**michael.farber@cs.uni-freiburg.de**

**Tel.: 0761/203-8133**