Datenbanken und Informationssysteme
Institut für Informatik
Dr. Michael Färber
michael.faerber@kit.edu

April 2018

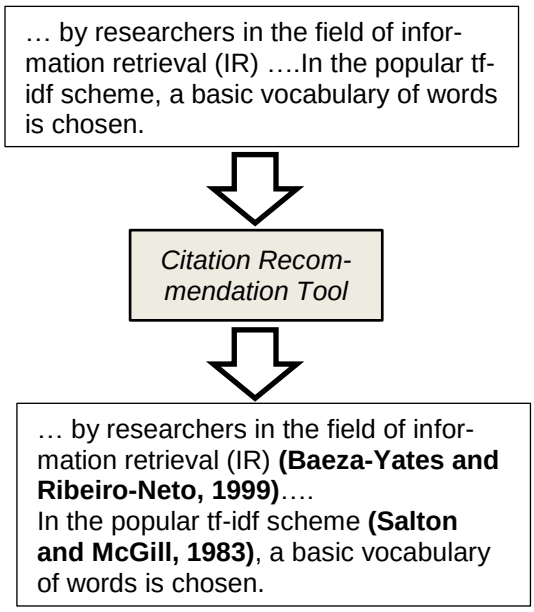
Call for Master Thesis

„No Time for Reading Literature? Developing a System for Automatically Adding Citations to Text“

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 developing and implementing an approach for adding (“annotating”) citations 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 (data sets with 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 (Java or Python or similar).
- Interest in applying methods of Text Mining/Natural Language Processing/Machine Learning.

Contact:
Michael Färber
michael.faerber@cs.uni-freiburg.de

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