

Call for Bachelor/Master Thesis „Can Citation Markers Be Words?” In [1] you may find an answer.“

[1] Your Thesis

What is the motivation and general goal?

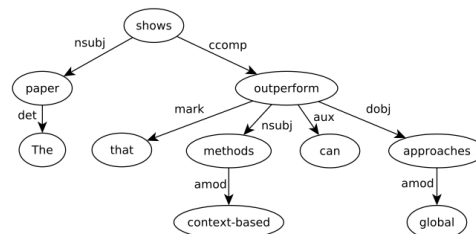
Current approaches to using machine learning methods on academic texts take a rather simplistic approach to dealing with in-text citations. Whatever is written somewhere close to a citation marker is treated as descriptive of the cited paper. This causes models to learn imprecise or even wrong representations of cited papers. To overcome this problem, it is important to be able to distinguish how a citation marker is used within a sentence. The goal of this thesis is to develop a model for this based on machine learning and natural language processing techniques.

What are the details?

Citation markers are sometimes used as an integral part of a sentence (“like a word”) and sometimes just shimmed in between the words. To detect how a citation marker is integrated into a sentence, the goal of the thesis is to develop a dependency parsing approach. That is, a model that can construct a tree like structure from a sentence showing how the individual words, including citation markers, depend on each other. The task therefore can be described as dependency parsing for the purpose of reference scope identification in citing sentences.

“While A et al. argue that X is well suited for Y [27], in [42] B et al. come to the conclusion that Z is a promising application area for X.”

“Other proposed pruning heuristics include pruning based on activations [3], redundancy [5], per-layer second derivatives [7], and energy/computation efficiency [9] (e.g., pruning convolutional filters [11-13] or channels [27]).”



What are the prerequisites?

- Interest in Data Mining, Machine Learning, and working with large data sets.
- No fear of actually implementing the methods you develop.