Karlsruher Institut für Technologie
Institut für Angewandte Informatik und
Formale Beschreibungsverfahren (AIFB)
Dr.-Ing. Michael Färber
michael.faerber@kit.edu

27.02.2019

Call for Master Thesis
„How Fair are Researchers? Analyzing Biases of Citations in Scientific Publications"

What is the topic?

Citing bias describes the phenomenon that authors (e.g., researchers writing a scientific publication) do not always cite objectively, but are biased towards different aspects.

- **Content of written text**: The actual content in the written text might be exposed to biases. For instance, scientific results (e.g., statistically non-significant studies are not published; specific outcomes from a study are reported, while others are not, etc.) are only reported selectively. It is also possible that authors tend to cite publications from specific disciplines and areas only (e.g., computer science, but not from psychology).
- **Authors of citeable publications**: It might be worthwhile to combat the unreasonable preference of self-citations and citations of papers authored by colleagues or friends.
- **Source of "citable" publications**: The "citable" publications' venues (e.g., especially prestigious venues) might be exposed to bias, too.

Note that citing is a difficult task in general and many citing biases might happen unwillingly and unknowingly. In this thesis, we will consider that by differentiating between explicitly and implicitly given citing biases.

The goal of this thesis is to **study different kinds of citing biases and to analyze how often or to which degree the citing biases appear** in reality. To that end, the data of the Microsoft Academic Graph (MAG) [1] will be analyzed. The MAG is a very large database containing information about publications, authors, venues, citations, keywords, etc. The MAG data and an initial list of publications concerning citing biases will be provided to the student. Based on the findings of the thesis, it should be possible to draw conclusions which kinds of citing biases are worth to be considered concerning automatic citation recommendation (i.e., how to minimize bias during the recommendation of citations).

[1] https://www.microsoft.com/en-us/research/project/microsoft-academic-graph/

Which prerequisites should you have?

- Interest in data mining and data analysis (also big data).
- Basic knowledge in statistics (e.g., correlation analysis).

Keywords: Scientific publications, citations, data mining, data analysis, bias detection.

Contact:
Dr.-Ing. Michael Färber
michael.faerber@kit.edu