

## Call for Bachelor/Master Thesis „A Systematic Review of Data Sets and Sources Providing Full Text Computer Science Publications“

### What is the general goal?

The goal is to systematically establish a comprehensive overview of data sets and general data sources which provide access to full text computer science publications. In a day and age where open access in science becomes increasingly prevalent and data more and more relevant, keeping an overview of available data resources in a given field is evermore important and challenging.

### What are the details?

With the growth of Open Access in academia and a need for automated processing of papers for knowledge extraction, an overview of machine readable scientific corpora is evermore important. Because of this, a large number of new data sets and data sources have been established within the last few years. The goal of the thesis is to identify *and inspect* such data sets and sources for the computer science domain and draw conclusions regarding use cases, mutual integratability, and possibilities for extension. For this, it will be necessary to not only read about, but also dive into the data.



### What are the prerequisites?

- Interest in data mining and working with large data sets.
- No fear of working on remote server infrastructure.