

Karlsruhe, 26. Februar 2019

## Visualisation of topic models

### Bachelor / Master Thesis (in German or English)

Topic models<sup>1</sup> are an increasingly popular text-mining tool. They rely on methods of unsupervised machine learning and natural language processing to identify topics within a body of text documents. Often, topic models are used to produce human-readable summaries of the content of a large collection of documents. However, to present the results of a topic model to the user in a clear and structured way can be challenging.

The aim of the proposed thesis is to explore the possibilities of presenting the results of a topic model to the user. Different visualisation techniques – possibly including interaction – can be used to facilitate the interpretation of topics. The use of additional information like topic labels, evaluation metrics (e.g. the exclusiveness of words in a topic), or context-specific metadata (e.g. publication year, classification, etc.) can further improve the visualisation.

#### Possible tasks:

- Give an overview over possible ways to visualise topic models.
- Use patent data provided by the chair to implement different visualisation techniques.
- Include information specific to the domain of patent data in the visualisation.
- Explore, implement, and test own ideas in the visualisation of topic models to facilitate the user's interpretation of the results.

#### What you should bring:

- Structured way of thinking and working
- An interest in machine learning and statistics
- Good programming skills esp. in Python

#### Are you interested?

- Contact: [david.baelz@kit.edu](mailto:david.baelz@kit.edu)

---

<sup>1</sup>See for example: <https://www.coursera.org/lecture/text-mining>.