

Call for Bachelor/Master Thesis “Extracting Facts from Text: Joint Entity and Relation Extraction”

(in English or German)

What is the topic?

The vast majority of information being made available online is contained in unstructured text (news/blog articles, scientific publications, etc.). The goal of information extraction is to develop methods to bring this information into a structured form (knowledge bases). Due to the complexity of this challenge, it is typically broken down into multiple steps, such as named entity recognition (NER), entity linking, relation extraction, event extraction, etc.

The goal of relation extraction is to detect relations between entities (such as a person or organization) mentioned in a text. Typically, this task is evaluated on text passages in which entities have been labeled manually. In joint entity and relation extraction this is not the case. Instead, both the entities as well as the relations between them have to be identified in an input text. This removes the necessity of labeling entities by hand and while this makes the task more difficult initially, multi-task setups can often improve performance on end-to-end tasks.

The focus of the proposed thesis is to compare existing approaches (such as [1],[2]) for joint entity and relation extraction and develop, implement, and test own approaches.

[1] <https://aclanthology.org/2021.findings-emnlp.204.pdf>

[2] <https://arxiv.org/pdf/2102.05980v1.pdf>

[3] <https://arxiv.org/pdf/2010.11304v3.pdf>

Which prerequisites should you have?

- Hands-on experience in machine learning, no fear to implement neural network models (under guidance of the supervisors).

John Stanistreet was an Australian politician. He was born in Bendigo to legal manager John Jepson Stanistreet and Maud McIlroy. (...4 sentences...) In 1955 John Stanistreet was elected to the Victorian Legislative Assembly as the Liberal and Country Party member for Bendigo. Stanistreet died in Bendigo in 1971.

Subject: John Stanistreet **Object:** Bendigo

Relation: place of birth; place of death

Figure 1: Example of text with annotated entities and relations.
Source: [2]

In case of good results, the findings of the thesis can be published as scientific paper together with the supervisor.

Starting date: As soon as possible.

Contact:

Dr. Michael Färber / Nicholas Popovic
michael.farber@kit.edu / popovic@kit.edu