Karlsruher Institut für Technologie
Institut für Angewandte Informatik und Formale Beschreibungsverfahren (AIFB)
Forschungsgruppe Web Science
M.Sc. Shuzhou Yuan
shuzhou.yuan@kit.edu

Call for Bachelor/Master Thesis:

"Text generation from graph structure for low-resource languages"

Background

Low-resource languages often suffer from a lack of available data and resources, which makes it difficult to develop effective natural language processing (NLP) systems for these languages. However, with the increasing availability of graph-structured data, there is an opportunity to leverage this structure to generate text in low-resource languages.

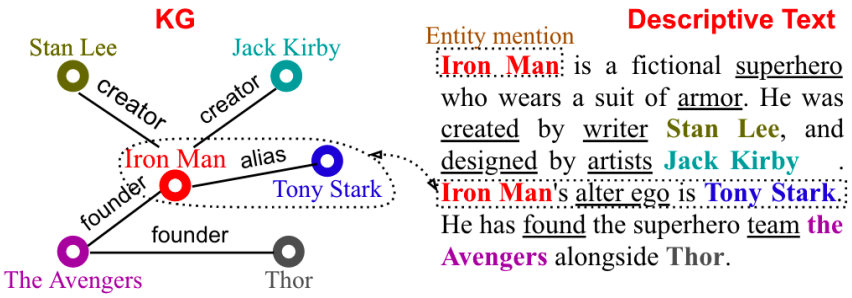

Figure 1: Text generation from knowledge graph in English

Previous work on graph-to-text generation has achieved outstanding performance using large language models [1]. However, the ability of these models to generate text in low-resource languages has not been thoroughly studied. The objective of this thesis is to explore the use of graph structure to generate text in low-resource languages. Specifically, the thesis will focus on developing a novel text generation approach that leverages graph-structured data, such as knowledge graphs, to generate high-quality text in low-resource languages.

Prerequisites

- Solid programming skills (e.g. Python).
- Strong interest in natural language processing and machine learning.
- Experience in pre-trained language models or HuggingFace library is a plus.

[1] https://aclanthology.org/2021.nlp4convai-1.20/

Please send your requests with a transcript of records and a short CV to:
Shuzhou Yuan
shuzhou.yuan@kit.edu