Incorporating Entity Typing and Location Information into Language Models to Improve Entity Linking

Knowledge graphs (KGs) provide structured semantic information for real-world entities and their relations, and can be accessed via SPARQL endpoints. Knowledge graphs are created using multiple sources including textual data, tabular data etc. The knowledge graphs are useful in entity linking across various text sources. The task of entity linking deals with aligning entities of the text to a knowledge graph which provides details about the mentions of the entities in the text. Consider the following example.

Berlin, Germany’s capital, dates to the 13th century. Reminders of the city’s turbulent 20th-century history include its Holocaust memorial and the Berlin Wall’s graffitied remains.

After entity linking with wikipedia will be,

Berlin, Germany’s capital, dates to the 13th century. Reminders of the city’s turbulent 20th-century history include its Holocaust memorial and the Berlin Wall’s graffitied remains.

There have been many approaches for entity linking, such as DBpedia Spotlight Daiber et al. [2013], using neural approaches Wu et al. [2020], Zhang et al. [2022], Logeswaran et al. [2019] and language models Broscheit [2019], Pörner et al. [2020].

Many aforementioned approaches, however, do not consider the entity types and location information. For example, there are many Holocaust memorials around the world. Using only string similarity can lead to inaccurate entity linking. To choose the right entity to link, we need the contextual information about not only the entity at hand but also the overall context of the text. Language models, such as Bert, Llama, GPT are useful in the task of entity linking since they consider the context of not only the entity but also the overall context of the given text. It is also beneficial to incorporate the location information to precisely link entities.

Another challenge is related to the entity type. In the following example, “the national football team of Germany won the world cup”, the entity is the German national football team and not the national football team and Germany as two different entities. Resolving these entity types and correctly linking them to the knowledge graph sources is a challenging task.

The goal of this thesis is to design an approach for entity linking between textual data and knowledge bases such as Wikidata, Dbpedia using entity types, location information and language models for accurate entity linking.

In this thesis, you will:

- Research on state-of-the-art methods for entity typing and entity linking. Design approach for the task in this thesis accordingly.
- Develop the method to automatically infer the links between text and knowledge bases.
- Evaluate the developed method on the benchmark datasets and compare its performance with state-of-the-art entity linking baselines.

Prerequisites:

- Knowledge of Semantic Web technologies and Knowledge Graphs, entity and relation linking e.g., obtained through successful completion of the module ”Advanced Methods of Information Retrieval” or other relevant modules and labs offered by the DSIS group.
• Theoretical and practical knowledge of machine learning along with language models
• Good programming skills

**Relevant Databases:** https://paperswithcode.com/datasets?task=entity-linking

**Key papers:**
• Fine-Grained Entity Typing for Domain Independent Entity Linking
• Improving Fine-grained Entity Typing with Entity Linking
• Do Judge an Entity by Its Name! Entity Typing Using Language Models
• Entity Linking with a Knowledge Base: Issues, Techniques, and Solutions
• Entity Hierarchy Embedding

**Contact:** Alishiba Dsouza

Please email dsouza@cs.uni-bonn.de along with your transcript and a link to GitHub (if available).

**References**


