M.Sc. thesis: Subtopic Relation Extraction for Knowledge Graph of Learning Concepts

Acquiring knowledge online has become one of the major ways in everyday learning. This includes learning in formal e-learning settings such as MOOC platforms, as well as in informal scenarios such as Web search, which is one of the most frequent online activities\(^1\). In this thesis, we focus on supporting human learning in Web search scenarios. Complex search tasks with a learning intent usually require multiple queries, and involve learning paths consisting of many different learning concepts (LCs), examples of LCs are shown in Fig. 1. DBpedia entities (or Wikipedia page titles) are often used as the source of LCs (e.g. [1]). To support such complex search tasks, it is crucial that the system is able to recommend appropriate resources and useful subsequent queries for each individual user.

Knowledge graph (KG), which is able to represent both entities (e.g. LCs) and between entity relations, has been widely used as a machine-readable knowledge source to support applications such as search engines, question answering systems and smart assistants etc. Representing LCs and their relations as a KG can enable the search system to understand learning tasks, and further improve functions such as query recommendation, query expansion and search result ranking.

In prior works, many relation types between LCs have been investigated, e.g. prerequisite, relatedness and subtopic relations [1-3]. Fig. 1 shows example relations between 6 LCs. Some prior works focused on mining such relations by training a supervised model based on a small annotated dataset, and applying it on Wikipedia entities to identify new relations [1-3].

![Fig. 1 Example of relations between LCs](image)

However, so far, there is no well-established method or resource that can provide large-scale and high-quality LC relation information that can be used to support learning in search. The task in this thesis is to build a KG consisting of LCs and sub-topic relations between LCs. Using annotated dataset and sub-topic relations in existing KGs, you will build machine learning/deep learning models for mining unknown sub-topic relations from KGs and from textual documents.

**Tasks**

- Research on state-of-the-art methods in entity relation extraction and KG completion [4, 5], find appropriate baselines and useful datasets.
- Analysis of the incompleteness of sub-topic relation in existing KGs (e.g. DBpedia\(^2\)).
- Design approach for the task in this project.
- Implement the designed approach and baseline methods.

---

\(^1\) [https://www.similarweb.com/top-websites/]

\(^2\) [https://www.dbpedia.org/]
- Evaluate the developed method on large-scale datasets, and compare its performance with state-of-the-art baselines.

**Prerequisites**

- Theoretical and practical knowledge of machine learning and deep learning.
- Good programming skills.
- Knowledge of Semantic Web technologies and Knowledge Graphs preferred, but required.

**Contact**

Dr. Ran Yu

Please email ran.yu@uni-bonn.de along with your transcript, and link to GitHub (if available).

**References**


