

Knowledge Graphs with Large Language Models

Assignment 4: Knowledge Graph Quality

Introduction

In this assignment you are asked to do a quality assessment of DBPedia. The DBPedia schema can be browsed [here](#) and the content can be accessed running SPARQL queries [here](#).

Task 1 (50%)

We want to assess the understandability of the DBPedia schema and one factor that plays a (negative) role is ambiguity. For that, you are asked to develop a term ambiguity detector and use it to identify ambiguous class and property names in the DBPedia schema. You can use an LLM for that purpose but, along with that, also use a dictionary-based approach (e.g. use [Wordnet](#) or [Wiktionary](#)). Comment on the effectiveness and efficiency of the dictionary-based approach vs the LLM approach.

Task 2 (50%)

As we discussed in class, abstract concepts are generally more difficult to accurately model than concrete ones. To see how well DBPedia does this kind of modeling, perform the following experiment:

- Consider the ESCO skill entities available in `esco_skills_en.csv` and find their equivalent entities (if any) in the DBPedia graph, as well as these entities' classes
- Use an LLM to judge if the entity-class pairs are accurate and describe your findings, including which DBPedia classes tend to have the most mistakes.

Deliverables

- An implementation of the task 1 and task 2 systems, either in the form of a Colab notebook or any other way you prefer, with clear instructions on how to run and evaluate them.
- A report describing the development and evaluation process and results for each task.