January, 2025

APM_5AI29_TP

Language Models and Structured Data

Mid-term Project Report

Acronym of the Team: AWESome
Name: Ardiansyah Nurgaha; Eddie Groh; Sriraam Appakutti; William Liaw.

This is the title of my project

Abstract

This midterm report presents a project combining PyKEEN and large language models (LLMs) to address knowledge graph completion, focusing on multi-class link prediction. By leveraging PyKEEN's graph embedding capabilities and LLMs' contextual understanding, the approach aims to predict missing links and classify their types, with applications such as drug repurposing through the inference of treats relationships in biomedical knowledge graphs.

Problem Statement

Explain which problem (and task) you're addressing. It should be framed as a problem (typically something desirable we want, or something undesirable we don't want, or something we don't understand, or something surprising).

[Ali et al., 2021]

Methodology

Neo4j Desktop Setup and Database Import

To facilitate knowledge graph completion, we utilize Neo4j Desktop for managing and querying the knowledge graph. This subsection provides a step-by-step guide to set up Neo4j Desktop and import a database dump file, as depicted in Figure 1. The specific dump used in this midterm report, derived from Hetionet, requires DBMS version 4.3 to ensure compatibility.

Upon successful import, the schema can be visualized using the following Neo4j query: `CALL db.schema.visualization()`. The resulting graph, illustrates key entities and relationships (e.g., *genes, compounds, diseases*) that form the basis for multi-class link prediction and knowledge graph completion in subsequent tasks.

Experimentation

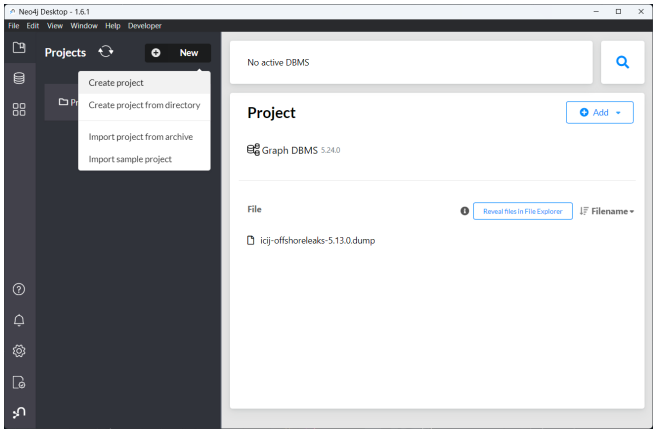
- Dataset & Dataset Statistics
- Experimental Results based on Evaluation Metrics
- Error Analysis

Discussion

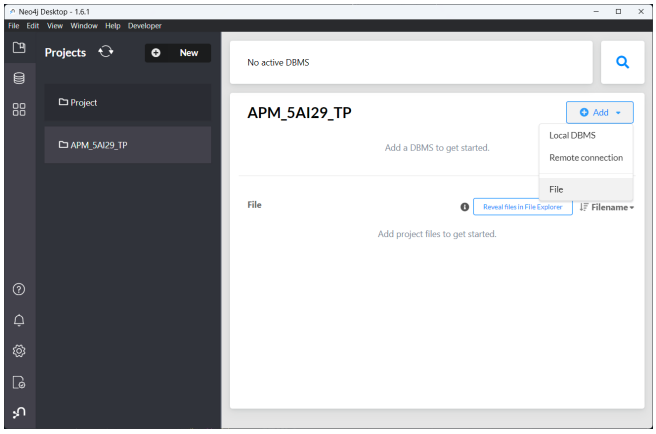
Comparison with expectations, limitations, lessons learned, and perspectives.

Bibliography

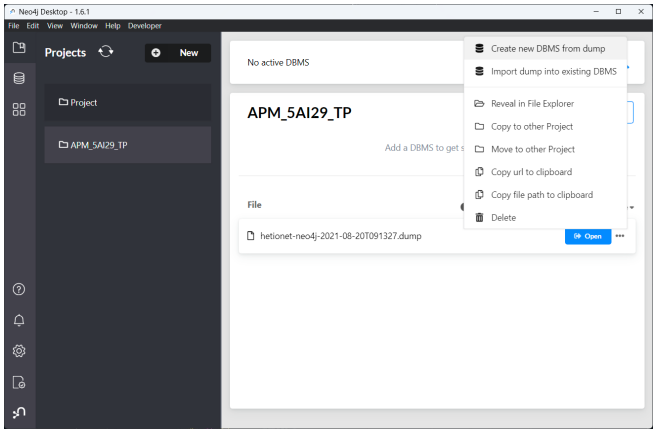
[Ali et al., 2021] Ali, M., Berrendorf, M., Hoyt, C. T., Vermue, L., Sharifzadeh, S., Tresp, V., and Lehmann, J. (2021). PyKEEN 1.0: A Python Library for Training and Evaluating Knowledge Graph Embeddings. *Journal of Machine Learning Research*, 22(82):1–6.



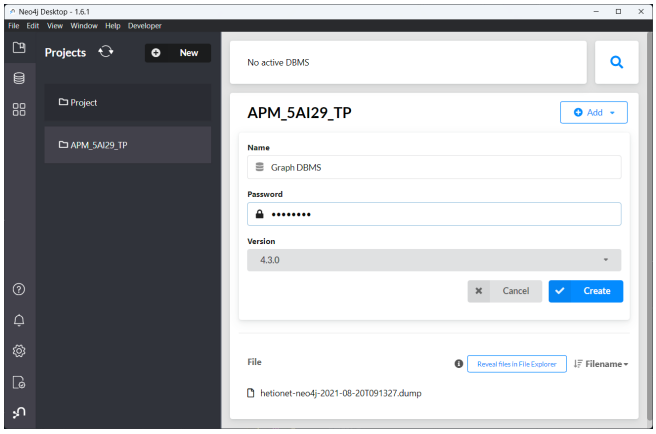
(a) Creating a project.



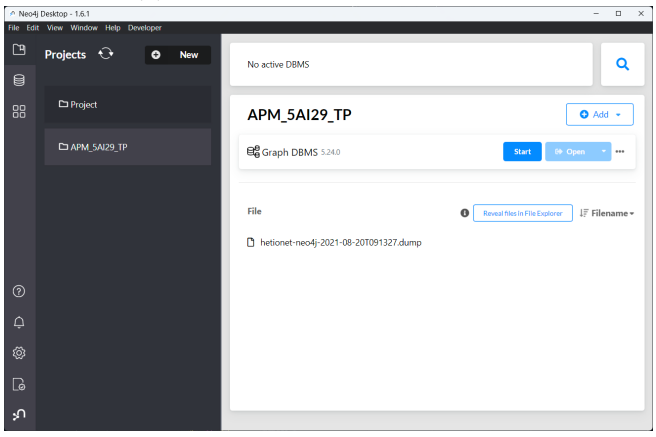
(b) Adding dump file.



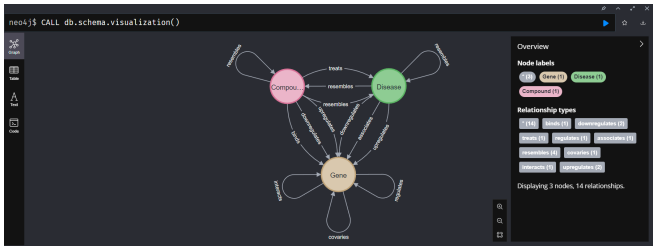
(c) Importing dump to DBMS.



(d) Configuring DBMS.



(e) Starting DBMS.



(f) Visualizing schema.

Figure 1: Steps for Neo4j Desktop Setup.