Building & Mining Knowledge Graphs

(KEN 4256)

Assignment 3

Multi-Hop Reasoning System on a Knowledge Graph

Due date: March 27, 2024 (upload on student portal before **23:59 CET**)

Description

Objective: Your task is to design, implement, and evaluate a multi-hop reasoning system that operates on a knowledge graph (KG). This assignment encourages creativity, critical thinking, and practical application of KG-based reasoning techniques you learned throughout the course.

Instructions:

1. KG selection
    a. Choose a KG that aligns with your interests or domain expertise. It could be a publicly available KG (e.g., Wikidata, DBpedia) or a custom-built one (including a subset of an existing KG).
    b. Justify your choice by how a set of multi-hop questions can be exactly answered by the KG.

2. System design and implementation
    a. Multi-hop reasoning
        i. Describe your approach to multi-hop reasoning. How will your system traverse the KG to infer complex relationships?
        ii. Implement the reasoning mechanism (e.g., rule-based, embedding-based, or neural network-based)
    b. Data processing
        i. Discuss any data processing steps taken.
        ii. Explain whether and how you handle missing data, noise, and inconsistent data.
    c. Evaluation metrics
        i. Define appropriate metrics to assess the system's performance (e.g., accuracy, precision, recall).
        ii. Implement evaluation.

3. Theoretical analysis
    a. Strengths & weaknesses

        i. Analyze the strengths and limitations of your reasoning system
        ii. Consider scalability, interpretability, and computational efficiency
    b. Zero-shot learning
        i. Discuss how your system handles zero-shot scenarios (i.e., unseen or rare associations)

4. Practical implementation
    a. Code repository
        i. Share your codebase (GitHub, GitLab, etc.) with clear documentation
        ii. Include instructions for running the system
    b. Demonstration
        i. Provide examples of multi-hop reasoning in action
        ii. Show how your system infers new facts or associations

5. Quality of writing
    a. Clarity & coherence
        i. Write a concise and well-structured report
        ii. Clearly explain your methodology, results, and insights
    b. Grammar & style
        i. Pay attention to grammar, spelling, and formatting
        ii. Use professional language

6. Creativity & freedom
    a. Appendix
        i. Include a dataset (or a subset) from your chosen KG
        ii. Describe any challenges faced during implementation
        iii. Reflect on lessons learned and future directions
    b. Exploration
        i. Encourage exploration beyond the basic requirements
        ii. Innovative ideas or extensions

Rubric for evaluation

1. Excellent (15 points)
    - The submission demonstrates both theoretical depth and practical implementation
    - Clear, concise, and insightful analysis
    - High-quality writing, engaging
    - Innovative approaches or extensions
2. Very good (12-14 points)
    - Solid theoretical understanding
    - Good practical implementation
    - Well-written report
    - Some exploration beyond the basics
3. Good (9-11 points)
    - Adequate theoretical analysis
    - Basic implementation
    - Acceptable writing quality

    - Minimal exploration
4. Satisfactory (6-8 points)
    - Meets minimum requirements
    - Limited depth
    - Writing needs improvement
5. Needs improvement (below 6 points)
    - Incomplete or flawed implementation
    - Poorly written report

Deliverables

- Report: *One student per group* will deliver a report, with a maximum of 5 pages, containing work distribution / individual-contribution section, indicating each member's contribution to the assignment. Include Disclosure and Reflection notes if AI was used for implementation.
- Code: A Python script containing implementation of your reasoning system **(in a single file, .py or .ipynb),** including minimal documentation to install any additional dependencies and run the whole pipeline
- Zip all deliverables to one file, name it as "Group x_Assignment 3"

Reading Materials

- Lecture 10 - RAG
- Lab10 - RAG over KG.ipynb
- All other lectures and labs provided in this course
- The FAIR Guiding Principles for scientific data management and stewardship: https://www.nature.com/articles/sdata201618

Questions and Comments:

Prof. Dr. Michel Dumontier, michel.dumontier@maastrichtuniversity.nl

Dr. Rena Yang, rena.yang@maastrichtuniversity.nl