

Assignment

ELL881/AIL821: Sem-I, 2024-25

Maximum Marks: 20

Due Date: November 14, Thursday, 11:59 pm

This assignment needs to be done INDIVIDUALLY. Do not plagiarize other's code/report. Include appropriate citations if you adapt code from any public sources or if you use methods proposed in any publicly available work.

The goal of this assignment is to implement a *knowledge graph-based retrieval-augmented generation (KG-RAG) model* for biomedical text generation tasks, based on the work by [Soman et al. \(2024\)](#), and further enhance their proposed method. Your tasks include – (1) setting up the existing KG-RAG framework, (2) exploring the potentials of some further improvement strategies, and (3) documenting your implementation and results comprehensively.

1 Tasks

I. Repository Setup and Reproduction

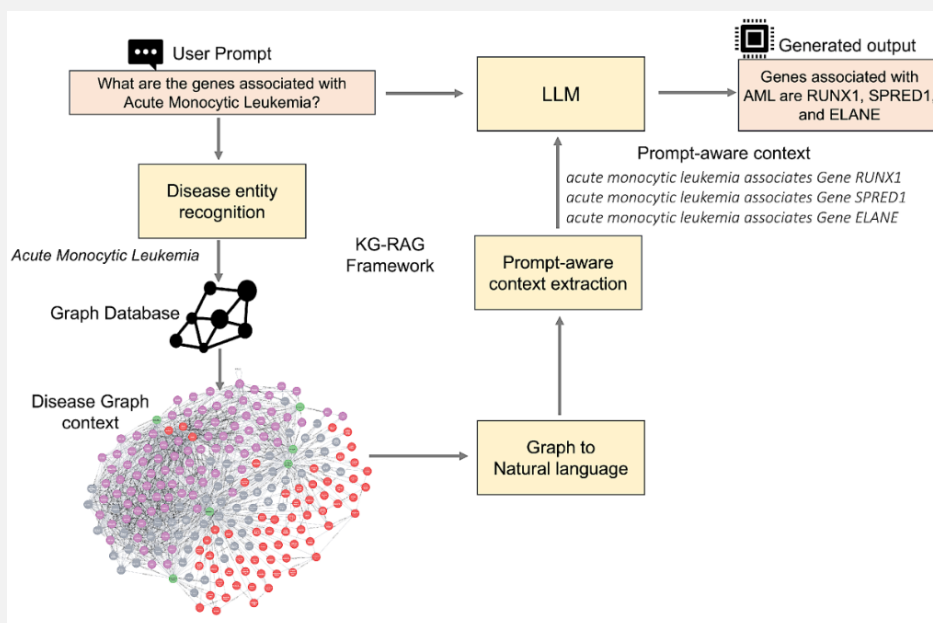


Figure 1: Framework for KG-RAG.

1. Access the repository at [LLM2401-Assignment Repository](#).
2. Follow the setup instructions to execute the script `run_gemini.sh`, in the following setting:
 - **Dataset:** [kg-rag/BiomixQA](#) (MCQs).
 - **LLM:** Gemini-1.5-flash for Disease Entity Extraction and Answer Generation. It is free to use at an acceptable speed (please set up your API key by referring to <https://makersuite.google.com/app/apikey>).

3. Obtain the model outputs and the accuracy score on the test set.

II. Implementation of Improvement Strategies

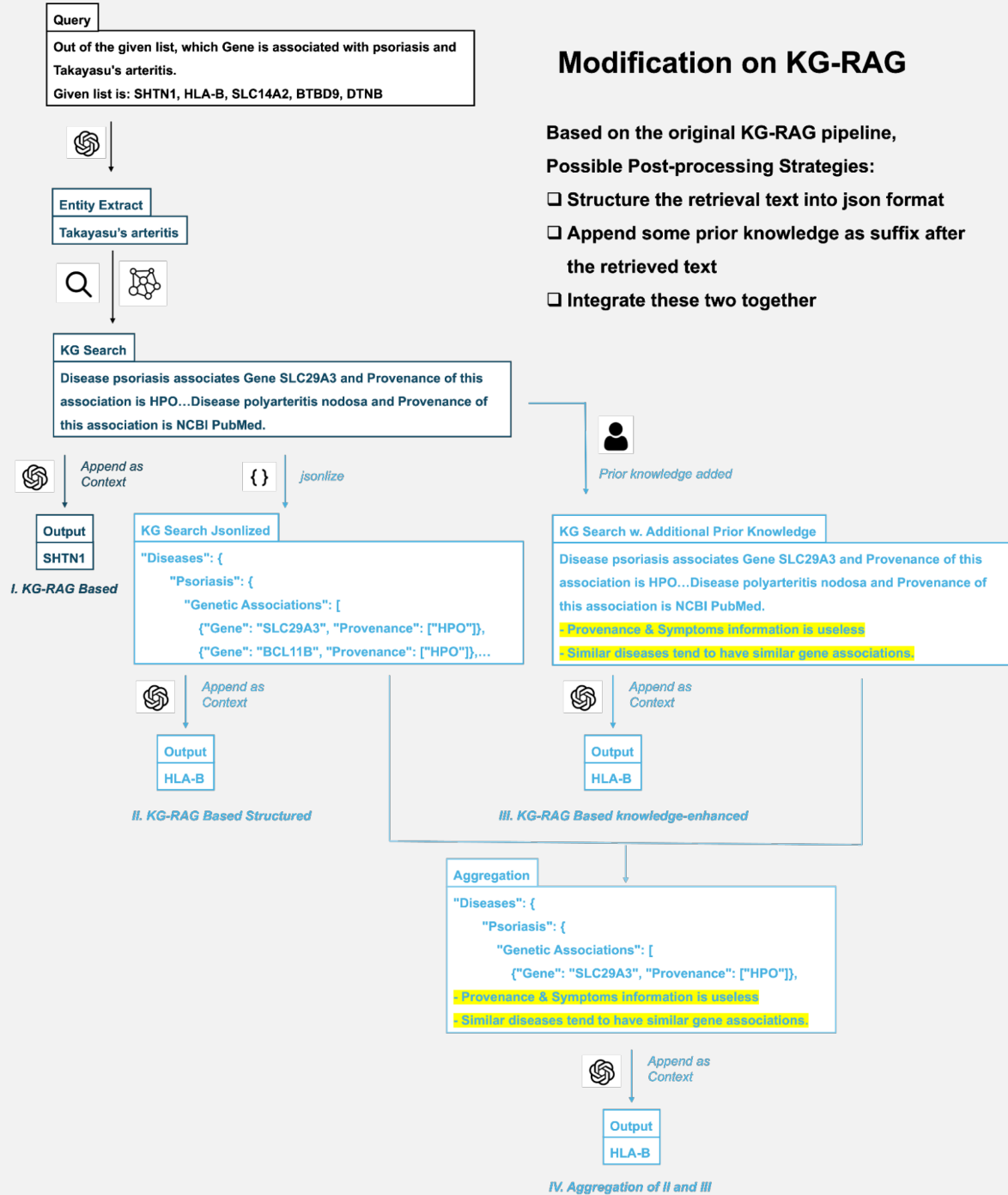


Figure 2: Three Improvement Strategies based on KG-RAG.

1. Implement the three improvement strategies, as shown in Figure 2, based on KG-RAG

to see if better performance can be achieved in the same task setting.

2. Obtain the outputs of three improvement strategies and the accuracy scores on the test set.

III. Report Writing

Prepare a comprehensive report detailing the motivation, implementation, and experiment results of the base model and three improved models.

2 Submission Guidelines

Submit the following as a `zip` file. The name of the file should be in the format:

`<Entry_number>_<Course_code>_A1.zip` (for example, `2023EEZ8463_ELL881_A1.zip`). **All submissions must be done through Moodle.** The assignment submission link will be available in the Moodle course page before the submission deadline.

- **Code:** Submit all modified code files, along with necessary instructions for running them (please include a README file for instructions on how to run your submitted code). Ensure that the code is well-documented and reproducible.
- **Model Outputs:** Four `csv` files, including the outputs of KG-RAG and that with the three improvement strategies.
- **Report:** The report should be no more than four pages in PDF format. It can include the following sections: Introduction, Method, Implementation, Experiment Result. You can also include any other section(s) based on how you want to present your work.

3 Evaluation Criteria

- **Accuracy of Reproduction (20%):** Clarity in replicating the results of KG-RAG.
- **Implementation of Improvements (60%):** Effectiveness of three modifications.
- **Quality of the Report (20%):** Clarity and completeness of the written report.

References

Soman, K., Rose, P. W., Morris, J. H., Akbas, R. E., Smith, B., Peetoom, B., Villouta-Reyes, C., Cerono, G., Shi, Y., Rizk-Jackson, A., Israni, S., Nelson, C. A., Huang, S., and Baranzini, S. E. (2024). Biomedical knowledge graph-optimized prompt generation for large language models. *arXiv: 2311.17330*.