Assignment 4: Naive RAG

Due: 2024-11-21 15:29 PM

Tasks

1. Try Naive RAG with Different Options: **8 pts**
2. Find Best Option: **2 pts**

Total 10 pts

Python Notebook for Assignment 4

- Download the Python Notebook from the eTL.
  - "**PE2024F_Assignment_4.ipynb**" file
- Or you can download the Notebook file from the Google Drive
  - https://drive.google.com/drive/folders/1beQM7D1gDv7iMb2TVNzIDvmxOKH5Y-0b?usp=drive_link

1. Try Naive RAG with Different Options

- You should apply different options for Naive RAG and compare the results with the default option
- I recommend you to test it with your own question
- Please include the **screenshots** of the results and **comparative analysis** in your report
- You will get **2 points for each comparison**

Assignment 4: Naive RAG

Source document

TODO: Try different chunking options

TODO: Try different splitter - split text by tokens

TODO: Try other vector indexing

TODO: Apply prompt compression

```
######## Please try with your own questions ########
rag_chain2.invoke("Tell me about the various chunking methods.")
```

2. **Find Best Option**

- You can combine options freely to find the best RAG option for the document
- Please include the **option combination** and **screenshot** of the result in your report
- You will get **2 points** for this task

Option combination example
- Chunking method: RecursiveCharacterTextSplitter with chunk size 2000, chunk_overlap=200
- Vector indexing: IndexFlatL2 with FAISS
- Prompt compression: True

Submission

- Due: 11/21 (Thu), 15:29 PM (right before class)
- After you have completed the assignment, you need to submit your **report** as a PDF via eTL assignment submission
    - Ex) 2024-00000.pdf
- We will calculate deducted score according to the following equation:
    - $(Original\_score) \times (1 - 0.1 \times ceil(delayed\_days))$
    - After 5 days (120 hours), there is no score.
    - Ex1) 50 hours late: 70% of the original score
    - Ex2) 115 hours late: 50% of the original score
    - Ex3) 120 hours late: 0% of the original score
- Concept level discussion is encouraged, but discussion of code/prompt directly related to assignments is not allowed. The assignments must be students' own work.