Chaklam Silpasuwanchai, Todsavad Tangtortan Updated 09 Mar 2025
AT82.03: Machine Learning

A6: Let's Talk with Yourself

In this assignment, apply RAG (Retrieval-Augmented Generation) techniques in Langchain framework to augment your chatbot that specializes in answering questions related to yourself, your documents, resume, and any other relevant information.

Note: You are ENCOURAGED to work with your friends, but DISCOURAGED to blindly copy other's work. Both parties will be given 0.

Note: Comments should be provided sufficiently so we know you understand. Failure to do so can raise suspicion of possible copying/plagiarism.

Note: You will be graded upon (1) documentation, (2) experiment, (3) implementation.

Note: This is a one-week assignment, but start early.

Deliverables: The GitHub link containing the jupyter notebook, a README.md of the GitHub, and the folder of your web application called 'app'.

Task 1. Source Discovery - Based on code-along/01-rag-langchain.ipynb, modify as follows:

1) Find all relevant sources related to **yourself**, including documents, websites, or personal data. Please list down the reference documents (1 point)
2) Design your **Prompt** for Chatbot to handle questions related to your personal information. Develop a model that can provide gentle and informative answers based on the designed template. (0.5 point)
3) Explore the use of other text-generation models or OPENAI models to enhance AI capabilities. (0.5 point)

Note: Groq also offers the llama3-70b model (generator model) with limited request capacity. For further exploration, refer to the following link[1].

Task 2. Analysis and Problem Solving

1) Provide a list of the retriever and generator models you have utilized. (0.25 point)
2) Analyze any issues related to the models providing unrelated information. (0.25 point)

Note: RAG utilizes two models: a **retriever** model and a **generator** model. Therefore, when performing your analysis, make sure to evaluate and analyze both models, not just one.

Task 3. Chatbot Development - Web Application Development - Develop a web application that demonstrates a chatbot.

1) The application should feature a chat interface with an input box where users can type messages.
2) Based on the user input, the model should generate coherent responses and also provide relevant source documents that support the generated response. For example, if the user types "How old are you?", the model might generate a concise summary along with links to related articles or documents. (0.5 point)

Note: You are encouraged to use any available resources related to your personal information, and ensure the chatbot provides accurate and relevant information.

Below are 10 questions your chatbot should be able to answer:

1) How old are you?
2) What is your highest level of education?
3) What major or field of study did you pursue during your education?
4) How many years of work experience do you have?
5) What type of work or industry have you been involved in?
6) Can you describe your current role or job responsibilities?
7) What are your core beliefs regarding the role of technology in shaping society?
8) How do you think cultural values should influence technological advancements?

[1] https://python.langchain.com/docs/integrations/chat/groq/

9) As a master's student, what is the most challenging aspect of your studies so far?
10) What specific research interests or academic goals do you hope to achieve during your time as a master's student?

Submission Instructions: For each question, your chatbot should generate a response. Please submit the question-answer pairs to your Github repository in the following JSON format:

```
[
  {
    "question": "How old are you?",
    "answer": "Your answer here"
  },
  {
    "question": "What is your highest level of education?",
    "answer": "Your answer here"
  },
  ...
]
```

Make sure that each question and corresponding answer is properly formatted in the JSON structure. This will be part of your deliverables. (0.5 point)

Good luck :-)