

Internship Assignment 2024 Coding Challenge

AI/ML Engineering Intern

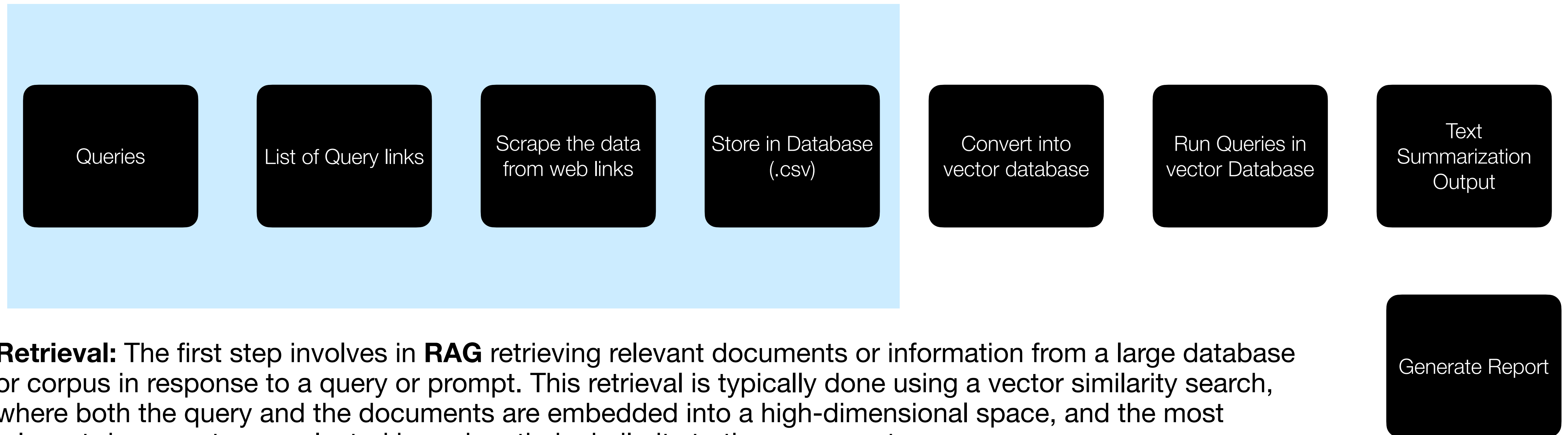
In the **Previous Assignment**, we provided you with a case study related to an electric vehicle (EV) manufacturing plant that wants to leverage AI, automation, and IoT to optimize its business operations. Your task was to conduct market research to identify target customers' needs, preferences, and purchasing patterns, analyze competitors, and provide insights into market trends and financial performance.

For **This Assignment**, we would like you to demonstrate how you would approach the task of gathering the required information for the case study using Python programming.

Specifically, we want you to write a Python script that searches the internet for information related to Canoo, a publicly traded company listed on NASDAQ (ticker symbol: GOEV).

Your script should retrieve data from various online sources and store it in a CSV file for further analysis.

Write a Python Code for **Retrieval**: The first step involves in **RAG**



Retrieval: The first step involves in **RAG** retrieving relevant documents or information from a large database or corpus in response to a query or prompt. This retrieval is typically done using a vector similarity search, where both the query and the documents are embedded into a high-dimensional space, and the most relevant documents are selected based on their similarity to the query vector.

RAG : Retrieval-Augmented Generation

Coding Task

Here are the specific tasks you should perform:

Your task is to write a script that will search the following queries using a internet search APIs, then scrape the data from the links and save in a structured tabular format in a *.csv file

1. Identify the industry in which Canoo operates, along with its size, growth rate, trends, and key players.
2. Analyze Canoo's main competitors, including their market share, products or services offered, pricing strategies, and marketing efforts.
3. Identify key trends in the market, including changes in consumer behavior, technological advancements, and shifts in the competitive landscape.
4. Gather information on Canoo's financial performance, including its revenue, profit margins, return on investment, and expense structure.

To help you get started, we have attached a sample report that provides an overview of the information we expect you to gather. Please note that this report is for reference only, and you are not expected to generate a similar report. Instead, we want to see how you would approach the task of extracting relevant data from various online sources using Python.

Coding Task

In your submission, please include the following:

1. A brief summary of the steps you took to complete the task, including any challenges you faced and how you overcame them.
2. A link to a GitHub repository containing your Python script and any necessary dependencies.
3. A sample output of the data retrieved from the internet, stored in a CSV file or other suitable format.

We look forward to reviewing your submission and assessing your skills in Python programming and data collection. Good luck!

For any question internship2024@lizmotors.com

References

Libraries

- scrapy
- selenium
- BeautifulSoup
- duckduckgo_search
- requests

Similar tool

- GPT researcher
- <https://github.com/assafelovic/gpt-researcher>

Link

- https://en.wikipedia.org/wiki/Large_language_model
- <https://www.ibm.com/docs/en/watsonx-as-a-service?topic=models-retrieval-augmented-generation>