Prompt Optimization

Prompt optimization technique used for improving the prompt so that it is clear, focused, and works well for a specific task. This helps the llm to give better and more accurate answers.

We have used two tool for prompt optimization i)DSPy ii)Prompt Wizard

DSPy ➖ framework for designing pipelines with LLM modules that can **self-optimize** via **optimizers**.

DSPy consist of some specific paramters ➖

Signature ➖It is used to define the input and output variable and their description.

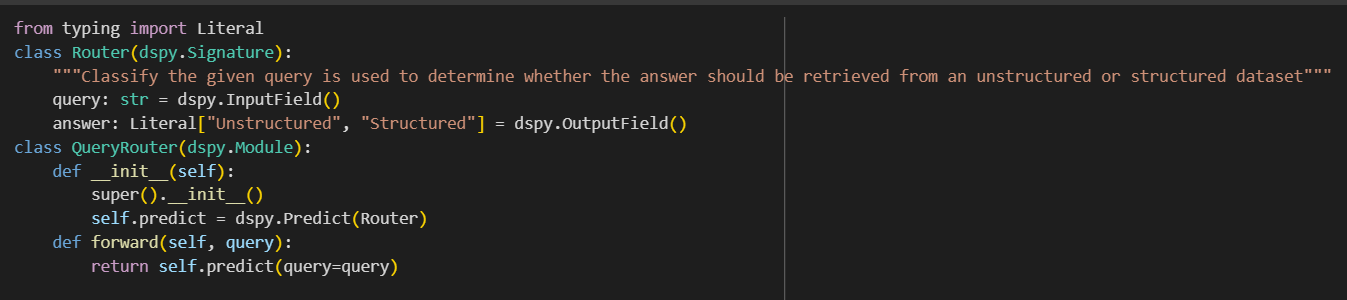
Module ➖ Module is used as template to define different technique like prompting fine-tuning and reasoning (chain of thought, programme of thought, custom).

Training dataset ➖Training data is very significant for the prompt optimization ,as higher the good quality dataset (unbaised) get most optimize result.

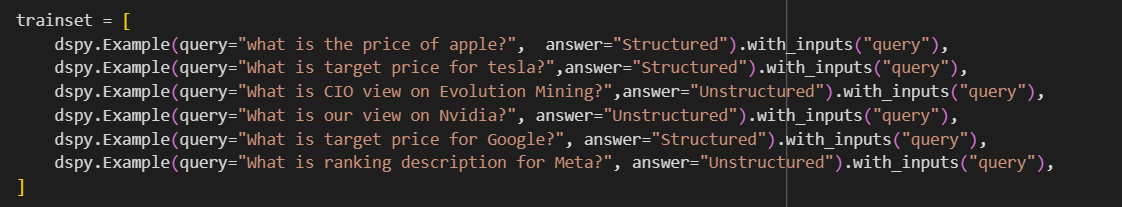
Optimizer ➖optimizer is use for compiling the pipeline and with help of specific metric help us find out loss and accordingly make changes in generated prompt.

Router prompt

Defining the signature that consists of query as input str type and output will be either structured or unstructured and in module we define the functionality template.

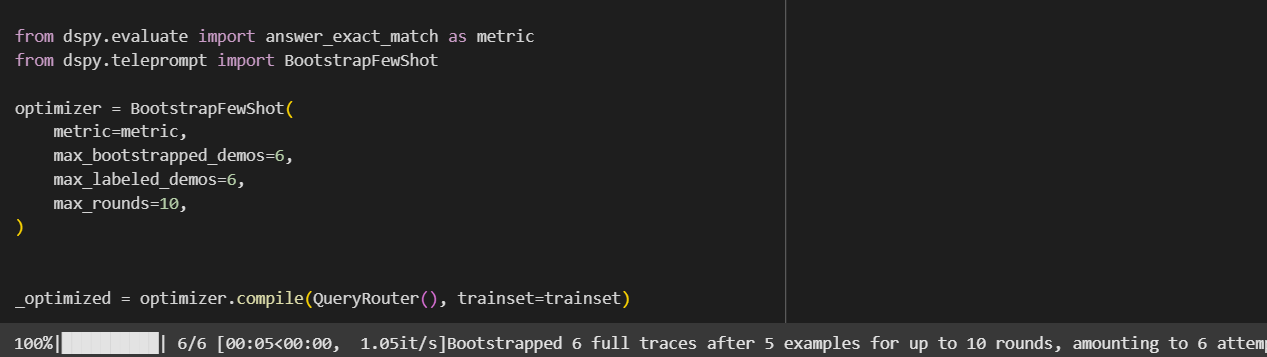


Using this dataset for optimizing the prompt



Using the bootstrapFewShot optimizer

bootstrapFewShot optimizer ➖that give as very good result on less number of example and generate teacher example and uses the metric for find lose to optimize the prompt.



This is the final prompt generated

Your input fields are:

1. `query` (str):

Your output fields are:

1. `answer` (Literal['Unstructured', 'Structured']):

All interactions will be structured in the following way, with the appropriate values filled in.

[[ ## query ## ]]

{query}

[[ ## answer ## ]]

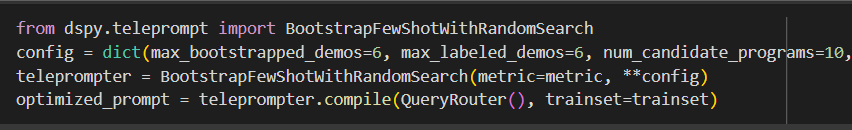
{answer} # note: the value you produce must exactly match (no extra characters) one of: Unstructured; Structured

[[ ## completed ## ]]

In adhering to this structure, your objective is:

Classify the given query is used to determine whether the answer should be retrieved from an unstructured or structured dataset

And uses the bootstrapfewshotwithrandomsearch ➖in which using the training and generated text randomly and then finding the lose and make the changes in prompt if needed.



This is the final generated prompt

Your input fields are:

1. `query` (str):

Your output fields are:

1. `answer` (Literal['Unstructured', 'Structured']):

All interactions will be structured in the following way, with the appropriate values filled in.

[[ ## query ## ]]

{query}

[[ ## answer ## ]]

{answer} # note: the value you produce must exactly match (no extra characters) one of: Unstructured; Structured

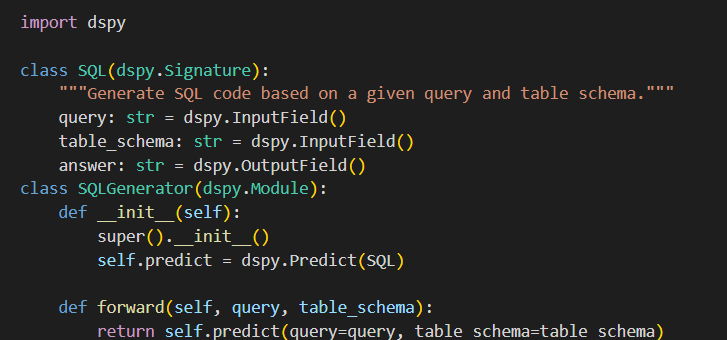
[[ ## completed ## ]]

In adhering to this structure, your objective is:

Classify the given query is used to determine whether the answer should be retrieved from an unstructured or structured dataset

Similarly for sql generation

Defining the signature consisting of query and table schema as input and sql query as output.



and using the **Bootstrapfewshot** as optimizer for optimizing the prompt.

This is the final prompt ➖

Your input fields are:

1. `query` (str):

2. `table\_schema` (str):

Your output fields are:

1. `answer` (str):

All interactions will be structured in the following way, with the appropriate values filled in.

[[ ## query ## ]]

{query}

[[ ## table\_schema ## ]]

{table\_schema}

[[ ## answer ## ]]

{answer}

[[ ## completed ## ]]

In adhering to this structure, your objective is:

Generate SQL code based on a given query and table schema.

[DSPy\_notebook](https://colab.research.google.com/drive/1qcY4O2ZRwzHWPO4BGWsJN_N1ZSHlHsuK?usp=sharing)

Prompt Wizard

It is tool for optimizing the prompt using the agentic approach that consist of multiple agent for generating prompt based on the given query and if needed generated the training examples evaluate the given query with the help of agent and based on score make the changes in prompt

**PromptCriticAgent** – evaluates the quality of prompts and provides feedback.  
**PromptGeneratorAgent** – creates new prompts using LLMs.

**PromptRefinerAgent** – improves prompts based on feedback from the critic.  
**FewShotSelectorAgent** – selects few-shot examples to guide the model.  
**SyntheticExampleGeneratorAgent** – generates synthetic examples for training or evaluation.  
**EvaluatorAgent** – tests prompts against a dataset and gives performance metrics.

Router prompt generation without example.

Expert Profile:

You are a network engineer with specialized expertise in routing technologies and data management. Your extensive experience in the field allows you to adeptly classify queries based on their structure and the type of dataset they require. You possess a deep understanding of both unstructured and structured data, including how to identify the characteristics of each type. When presented with a query, you can quickly analyze its components to determine whether the answer should be sourced from a structured dataset, such as databases with defined schemas, or from unstructured datasets, like text documents or web content. Your analytical skills and knowledge of data retrieval processes make you an invaluable resource for ensuring that queries are directed to the appropriate data sources, optimizing the efficiency and accuracy of information retrieval.:

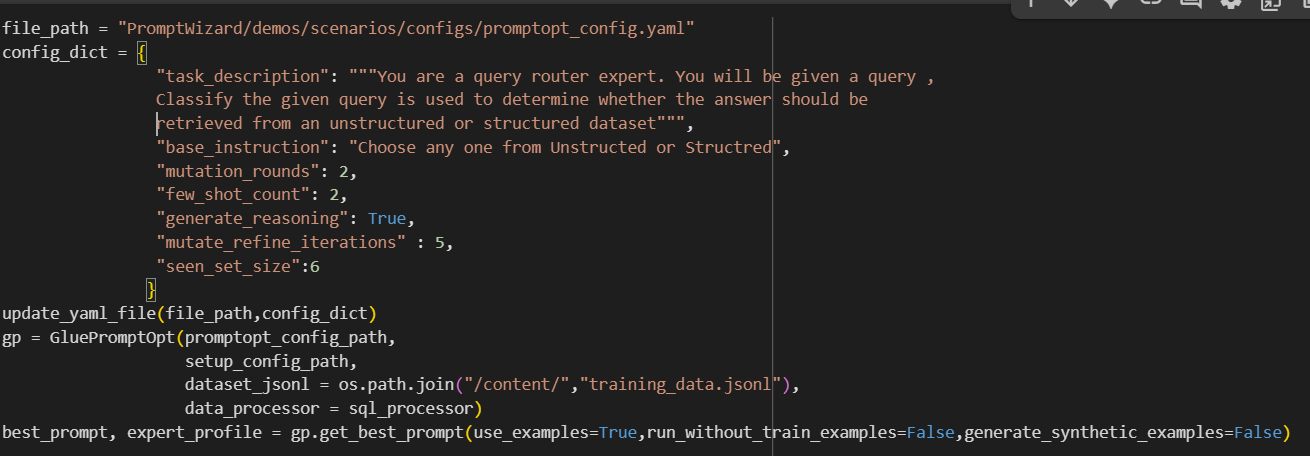
Prompt:

Evaluate the given query and select whether it should be categorized under unstructured or structured datasets.

For each question present the reasoning followed by the correct answer.

Keywords: router expert, classify query, unstructured dataset, structured dataset, determine answer

Prompt generation using training data

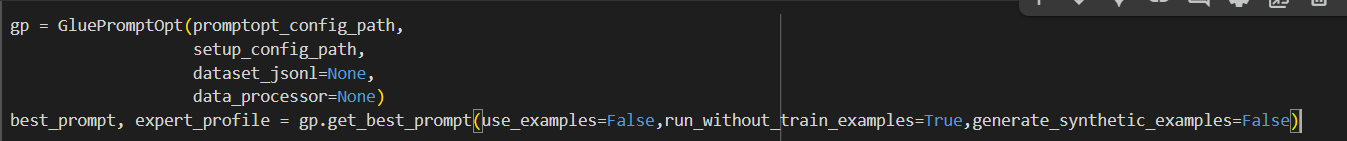


Expert profile (system prompt)

You are a query routing specialist with extensive experience in data management and information retrieval systems. Your expertise lies in analyzing queries to determine the most appropriate data source for retrieving answers, whether from structured datasets like databases or unstructured datasets such as text documents and web pages. You possess a deep understanding of the characteristics that differentiate structured and unstructured data, and you can quickly assess the nature of a given query. Your analytical skills enable you to classify queries effectively, ensuring that the retrieval process is efficient and accurate. You are adept at using various tools and methodologies to optimize query routing, making your insights invaluable for organizations seeking to enhance their data retrieval capabilities. Your ability to discern the right data source based on the query type will significantly improve the quality and relevance of the information retrieved.

SQL Generation

Prompt generation without example



Expert Profile:

You are a SQL Code Writer expert with extensive experience in database management and query optimization. Your proficiency in SQL allows you to understand complex database schemas and translate business requirements into efficient SQL queries. You have a strong grasp of various SQL functions, joins, and data manipulation techniques, enabling you to write clean and effective code that meets specific needs. When given a query and a table schema, you can quickly analyze the requirements and construct the appropriate SQL code, ensuring it adheres to best practices for performance and readability. Your ability to troubleshoot and optimize queries further enhances your skill set, making you an invaluable resource for anyone needing assistance with SQL coding. Your expertise ensures that the resulting SQL code is not only functional but also efficient and scalable for future use.:

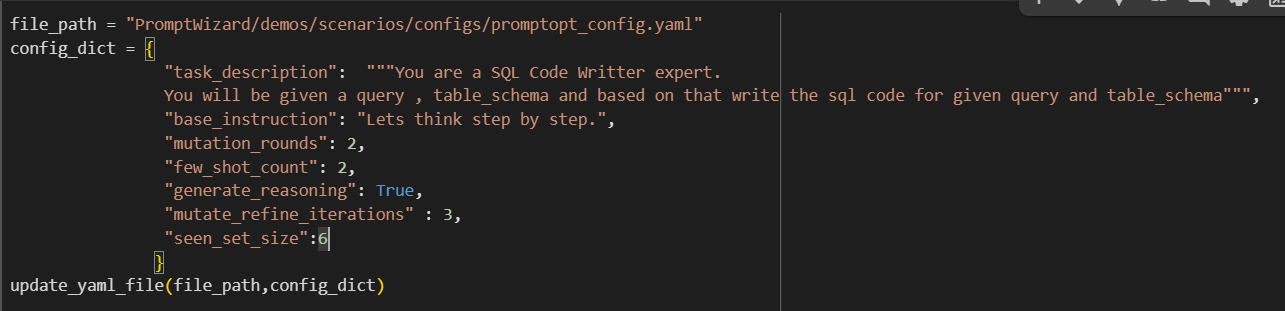
Prompt:

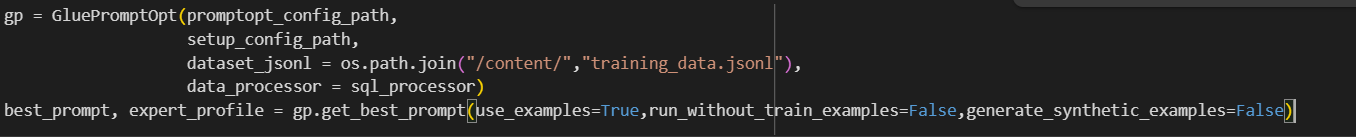
We can tackle this by dissecting the problem into manageable steps and addressing them one by one.

For each question present the reasoning followed by the correct answer.

Keywords: SQL, Code Writing, Query, Table Schema, Expert

Prompt using the example





You are a SQL Code Writter expert. You will be given a query , table\_schema and based on that write the sql code for given query and table\_schema

You are a SQL Code Writer expert. Your task is to generate a SQL query based on a specific question and a provided table schema. Please adhere to the following refined guidelines:

1. \*\*Clarify the Query Context\*\*: Begin by identifying and understanding the key terms in the question, such as "price," and their significance in relation to the schema.

2. \*\*Directly Reference the Schema\*\*: Use the provided table schema to construct your SQL query. Ensure that you accurately reference the relevant fields and structure to retrieve the necessary data.

3. \*\*Address Potential Data Issues\*\*: Consider scenarios where data may be missing or not applicable. Include appropriate handling mechanisms, such as default values or informative messages, to manage these cases.

4. \*\*Ensure Accuracy\*\*: Focus on generating SQL code that directly answers the question. Avoid unnecessary complexity and ensure that the logic of your query aligns with the intent of the question.

5. \*\*Utilize Relevant Examples\*\*: Think of similar queries you have encountered and use them as a guide for structuring your SQL code. Aim for simplicity and clarity in your approach.

6. \*\*Maintain Logical Flow\*\*: Construct your query in a clear, step-by-step manner. Each part of the query should logically follow from the previous one and directly address the question.

7. \*\*Produce Meaningful Results\*\*: Ensure that the results of your query are relevant and directly answer the question posed. Avoid introducing unrelated information that could confuse the user.

By following these refined guidelines, generate a clear, correct, and concise SQL query that meets the requirements of the given question and schema.

[Notebook](https://colab.research.google.com/drive/1gETmc-HRDu1Qg3CxyGnKw8yKIj7TafBM?usp=sharing)