GenAI based RAG using RapidMiner

Operators Used:
1. Execute Python
2. Retrieve
3. Store
4. Documents to Data
5. Read Document
6. Open AI Prompt

Steps:
1. Install the Python scripting and Text processing packages from the marketplace extensions.
2. Configure the Python environment path
3. Use an Execute Python operator with customized script to download and write the required file to the rapidminer database
4. Call the Open AI prompt operator and edit the main python script to include context and enable chat completion using langchain packages.

Demos:Prompts 1-5 with retrieved answers are as follows.

Views: Design **Results** Turbo Prep Auto Model Interactive Analysis

Find data,

Result History   ExampleSet (OpenAI Add Column)

Open in Turbo Prep Auto Model Interactive Analysis

Filter (1 / 1 examples): all

Data

Row No.	input	new_col
1	What is the do...	The document...

The document is about the future of petrochemicals and their impact
on the environment, energy system,
and sustainability. It discusses the current state of the chemical sector,
potential futures for the
industry, and policy recommendations for transitioning towards a more
sustainable petrochemical industry.

Press "F3" for focus.

Statistics

Visualizations

Annotations

ExampleSet (1 example,0 special attributes,2 regular attributes)

Views: **Design** Results Turbo Prep Auto Model Interactive Analysis

Find data, operators...etc   All Studio

Repository

Import Data

- Training Resources (connected)
- Community Samples (connected)
- Samples
- Local Repository (Local)
  - Connections
  - data
    - new_pdf (1/18/24 3:02 PM – 807 kB)
    - sample (1/17/24 5:52 PM – 1 kB)
  - processes
  - downloaded_file.pdf (1/18/24 11:33 AM –)

Process

Process ▸

Set Org ID   Set Prompt   Multiply

Set API Key   Set Name   Retrieve

OpenAI Add Column   Metadata Helper

Read Document   Documents to Data   Store

Operators

store

- Data Access (1)
  - Store
- Utility (1)
  - Process Control (1)
    - Remember
  - Extensions (1)
    - Text Processing (1)
      - Process Documents from Mail Store

No results were found.

Recommended Operators

| Category Parameter Macro 75% | Execute Script 68% | Integer Parameter Macro 57% |

Parameters

Set Prompt (Set Macro)

macro	prompt
value	What is the impact of petrochemicals on the environment?

Help

Set Macro
RapidMiner Studio Core

Tags: Process, Variable, Setting, Macros

Synopsis

This operator can be used to define a macro which can be used by %
{macro_name} in parameter values of succeeding operators of the
current process. The macro value will NOT be derived from any
ExampleSet. A macro can be considered as a value that can be used by
all operators of the current process that come after the macro has been

Views: Design | Results | Turbo Prep | Auto Model | Interactive Analysis

Find data, operators...etc

Result History | ExampleSet (OpenAI Add Column)

Open in | Turbo Prep | Auto Model | Interactive Analysis

Filter (1 / 1 examples): all

Data

Row No.	input	new_col
1	What is the im...	Petrochemical...

Statistics

Petrochemicals have a significant impact on the environment,
including contributing to climate change,
air pollution, and water pollution. However, solutions to mitigate these
impacts are achievable and cost-effective.

Visualizations

Press "F3" for focus.

Annotations

ExampleSet (1 example,0 special attributes,2 regular attributes)

Repository

Import Data

- Training Resour...
- Community Sam...
- Samples
- Local Repository
  - Connections
  - data
    - new_pdf (
    - sample (1/
  - processes
  - downloaded_fi
  - openai_demo
  - rag_based_op
  - translation_de
- Temporary Repos
- DB (Legacy)

Repository

Import Data

- Training Resources (connected)
- Community Samples (connected)
- Samples
- Local Repository (Local)
  - Connections
  - data
    - new_pdf (1/18/24 3:02 PM – 807 kB)
    - sample (1/17/24 5:52 PM – 1 kB)
  - processes
  - downloaded_file.pdf (1/18/24 11:33 AM –

Operators

store

- Data Access (1)
  - Store
- Utility (1)
- Process Control (1)
  - Remember
- Extensions (1)
- Text Processing (1)
  - Process Documents from Mail Store

No results were found.

Process

Process ▶

Set Org ID | **Set Prompt** | **Multiply** | **OpenAI Add Column** | **Metadata Helper**

Set API Key | **Set Name** | **Retrieve**

Read Document | **Documents to Data** | **Store**

Recommended Operators

| Category Parameter Macro 75% | Execute Script 68% | Integer Parameter Macro 57% |

Parameters

Set Prompt (Set Macro)

macro	prompt
value	What are products that can be produced using petrochemicals?

Help

Set Macro
RapidMiner Studio Core

Tags: Process, Variable, Setting, Macros

Synopsis

This operator can be used to define a macro which can be used by %
{macro_name} in parameter values of succeeding operators of the
current process. The macro value will NOT be derived from any
ExampleSet. A macro can be considered as a value that can be used by
all operators of the current process that come after the macro has been

Result History

ExampleSet (OpenAI Add Column) ✕

Open in | Turbo Prep | Auto Model | Interactive Analysis

Filter (1 / 1 examples): all

Row No.	input	new_col
1	What are prod...	Products that ...

Products that can be produced using petrochemicals include ethylene,
propylene, ammonia, methanol, toothbrushes,
carrier bags, food packaging,

Press "F3" for focus.

Data

Statistics

Visualizations

Annotations

ExampleSet (1 example,0 special attributes,2 regular attributes)

Repository ✕

⊕ Import Data

▶ 🖥 Training Resource
▶ 🖥 Community Samp
▶ 🖥 Samples
▼ 🖥 Local Repository
  ▶ 🔌 Connections
  ▼ 📁 data
      📄 new_pdf (1
      📄 sample (1/
  ▶ 📁 processes
      📄 downloaded_fil
      ⚙ openai_demo (
      ⚙ rag_based_ope
      ⚙ translation_de
▶ 🖥 Temporary Repos
▶ 🖥 DB (Legacy)

Process

Process ▸

Set Org ID | Set Prompt | Multiply
Set API Key | Set Name | Retrieve
OpenAI Add Column | Metadata Helper
Read Document | Documents to Data | Store

Process 7 s ↳ [1] OpenAI Add Column 7 s

Recommended Operators ⓘ

Category Parameter Macro 75% | Execute Script 68% | Integer Parameter Macro 57%

Parameters

Set Prompt (Set Macro) | ✕

| macro | prompt |
| value | State the importance of recycling petrochemical products. |

Help

Set Macro
RapidMiner Studio Core

Tags: Process, Variable, Setting, Macros

Synopsis

This operator can be used to define a macro which can be used by %{macro_name} in parameter values of succeeding operators of the current process. The macro value will NOT be derived from any ExampleSet. A macro can be considered as a value that can be used by all operators of the current process that come after the macro has been

Views: Design | Results | Turbo Prep | Auto Model | Interactive Analysis

Find data, operators...etc | All S

Result History | ExampleSet (OpenAI Add Column) | ✕

Open in | Turbo Prep | Auto Model | Interactive Analysis

Filter (1 / 1 examples): all

Row No.	input	new_col
1	State the impo...	Recycling petr...

Recycling petrochemical products is important because it can
eliminate demand for ethylene, save oil-equivalent,
reduce ocean-bound plastic waste, and is a more sustainable option
compared to incineration and landfilling.

Press "F3" for focus.

- Data
- Statistics
- Visualizations
- Annotations

ExampleSet (1 example,0 special attributes,2 regular attributes)

Views: Design | **Results** | Turbo Prep | Auto Model | Interactive Analysis

Find data, operat...

Result History | **ExampleSet (OpenAI Add Column)** ✕

Open in | Turbo Prep | Auto Model | Interactive Analysis

Filter (1 / 1 examples): all

Data

Row No.	input	new_col
1	What are the p...	The procedur...

The procedures followed during manufacturing petrochemical products include steam reforming of natural gas, synthesis steps such as Haber Bosch synthesis and methanol synthesis, and the use of oil feedstocks. China also uses coal as a feedstock, but this is a highly energy-intensive process. Some companies are also exploring downstream opportunities and integrated petrochemical units.

Press "F3" for focus.

Statistics

Visualizations

Annotations

ExampleSet (1 example 0 special attributes 2 regular attributes)