Informatica® Cloud Application Integration
August 2024

Simple React AI Agent with Search Tools

Informatica Cloud Application Integration Simple React AI Agent with Search Tools
August 2024

Publication Date: 2024-08-27

Table of Contents

Preface

Use *Simple React AI Agent with Search Tools* to learn how to use Google Search to gather information, perform tasks, and answer user queries. This guide assumes that you have an understanding of the Google API Connector and Gemini Connector concepts.

CHAPTER 1

Introduction to Simple React AI Agent with Search Tools recipe

The Simple React AI Agent with Search Tools recipe is based on REST and SOAP APIs. The recipe shows you how to build a simple React Agent with a search tool, Google Search, using the Serp API. The React AI agent logic alternates between generating thoughts and performing task-specific actions dynamically.

The Gemini Large Language Model (LLM) follows a step-by-step problem-solving approach, utilizing the search tool to gather information and perform tasks to answer user queries. The process orchestrates the whole workflow based on the LLM response.

The process accepts a user query. It passes the query to the LLM with additional system instructions to respond in a format that includes a question, thought, action, action input, observation, final thought, and final answer.

After receiving the response from the LLM, the process checks for the presence of the action and action input fields. If these fields are present, a request for the SearchTool service is formulated. Otherwise, it creates the final answer.

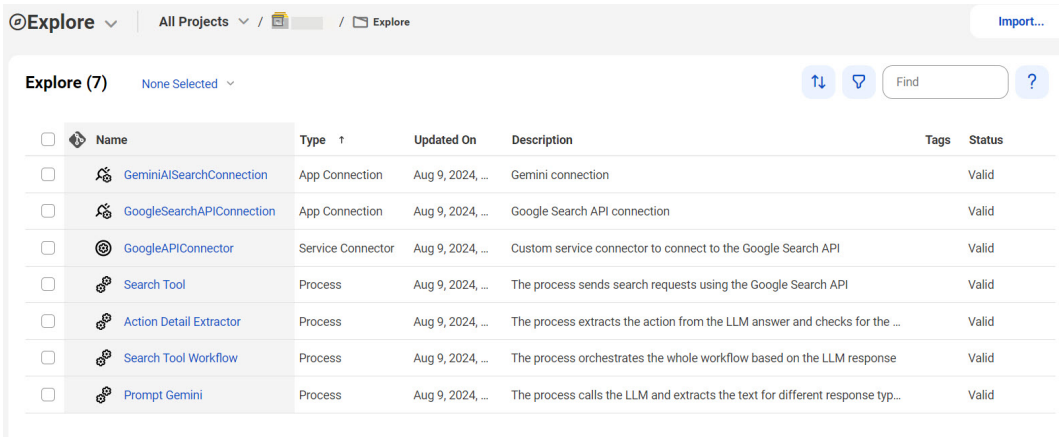
If the final answer is not set, a request is sent to the SearchTool service. The parent process is called with new parameters, such as user query and Agent_Scratchpad. Agent_Scratchpad is a template that contains fields. For example, Action: Action Input: Observation.

The process repeats until all the actions from the initial LLM response are processed or the number of repetitions does not exceed the specified limit. When the process execution time approaches 60 seconds, the process provides a final response. If the process ends prematurely, the final response might be inaccurate.

Simple React AI Agent with Search Tools contents

The Simple React AI Agent with Search Tools recipe contains app connections, a service connector, and processes.

The following image shows the assets that the Simple React AI Agent with Search Tools recipe package contains:



Simple React AI Agent with Search Tools recipe assets

The following table lists the assets that the Simple React AI Agent with Search Tools recipe package contains:

Asset Name	Asset Type	Description
GeminiAISearchConnection	App Connection	Gemini connection.
GoogleSearchAPIConnection	App Connection	Google Search API connection.
GoogleAPIConnector	Service Connector	Custom service connector to connect to the Google Search API.
Search Tool	Process	Sends search requests using the Google Search API.
Action Detail Extractor	Process	Extracts the action from the LLM answer, checks for the final answer, verifies that the execution time does not exceed 50 seconds, and monitors the number of times the process is called.
Search Tool Workflow	Process	Orchestrates the whole workflow based on the LLM response.
Prompt Gemini	Process	Calls the LLM and extracts the text for different response types, such as Python, JSON, SQL, and SOQL.

CHAPTER 2

Using the Simple React AI Agent with Search Tools recipe

To use the Simple React AI Agent with Search Tools recipe, you must perform the following steps manually:

Step 1: Copy and access the recipe

Step 2: Publish the GoogleAPIConnector service connector

Step 3: Configure and publish the GeminiAISearchConnection connection

Step 4: Configure and publish the GoogleSearchAPIConnection connection

Step 5: Configure and publish the processes

Step 6: Invoke the process

Copying and accessing the recipe

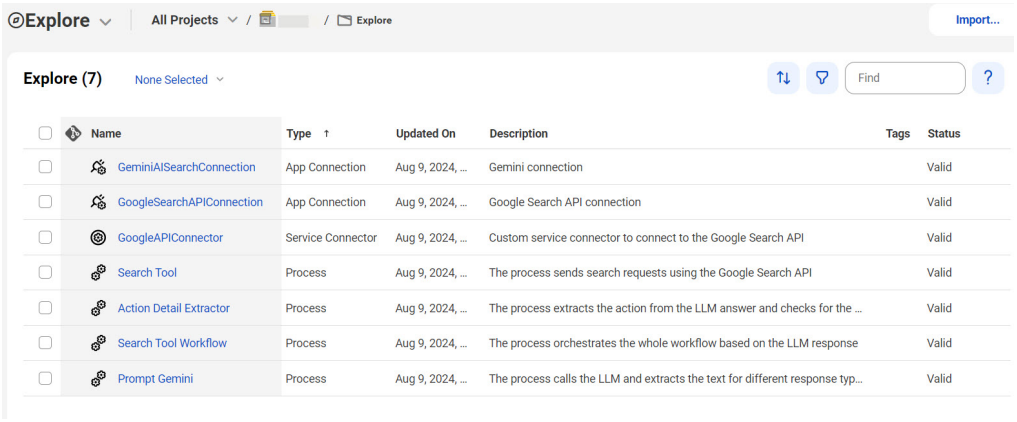
Copy the pre-configured assets in the recipe to a separate project or folder.

1. Open the **Simple React AI Agent with Search Tools** recipe and click **Use**.
2. Select the location where you want to copy the recipe, and then click **Continue**.
3. In the **Copying the recipe** dialog box, click **OK**.

   It might take some time for the recipe to get copied. You will receive a notification when the recipe is ready for use.
4. After the recipe is copied, click **Explore** to access the recipe content.

5. Navigate to the project or folder where you copied the recipe or enter the recipe name in the **Find** box.

   All the assets in the recipe are displayed as shown in the following image:



Publishing the GoogleAPIConnector service connector

You can publish the **GoogleAPIConnector** service connector from the assets page.

To publish the **GoogleAPIConnector** service connector, click **Actions** in the row that contains the service connector and select **Publish**.

Configuring and publishing the GeminiAISearchConnection connection

Configure the API key in the GeminiAISearchConnection connection, and then publish the connection.

1. Open the **GeminiAISearchConnection** connection.
2. In the **Connection Properties** section, enter the API key in the **API_Key** property. The **API_Key** property authenticates Gemini connection requests.
3. Save and publish the connection.

Configuring and publishing the GoogleSearchAPIConnection connection

Configure the API key in the GoogleSearchAPIConnection connection, and then publish the connection.

1. Open the **GoogleSearchAPIConnection** connection.

2. In the **Connection Properties** section, enter the API key in the **API_KEY** property. The **API_Key** property authenticates Google Search API connection requests.
3. Save and publish the connection.

Note: You can get the API key for the Google search tools from https://serpapi.com/.

Configuring and publishing the processes

To configure and publish the processes, perform the following steps:

1. To publish the **Search Tool** process, click **Actions** in the row that contains the process and select **Publish**.
2. To publish the **Action Detail Extractor** process, click **Actions** in the row that contains the process and select **Publish**.
3. Open the **Prompt Gemini** process.
4. On the **Temp Fields** tab of the Start step, the **Model_LLM** field is set to **gemini-1.5 pro** by default. You can optionally edit the model version. For information about changing the model version, see the Gemini documentation.
5. Save and publish the process.
6. Open the **Search Tool Workflow** process.
7. On the **Temp Fields** tab of the Start step, the **Loop_Limit** field is set to **5** by default. You can optionally edit the limit. This field defines the number of times the Search Tool subprocess can be called.

   Note: You can adjust the number of repeat queries in the Search Tool to get new information and better results. However, this will increase the number of tokens used.
8. On the **Assignment** tab of the **Create Prompt to LLM** step, the **Prompt** field is configured to contain the request template for the prompt to LLM. Optionally, you can configure the LLM behavior in the **Prompt_Request** field using the Expression Editor, as shown in the following sample code:

```
For Prompt_Request:
<GenerateContentRequestBody>
   <contents>
      <parts>
         <text>{$temp.Prompt}</text>
      </parts>
      <role>user</role>
   </contents>
    <generationConfig>
      <maxOutputTokens>2048</maxOutputTokens>
      <temperature>0</temperature>
      <topP>1</topP>
      <topK>250</topK>
   </generationConfig>
</GenerateContentRequestBody>
```

For the **Prompt_Request** field, enter values for the following properties:

Property	Description
maxOutputTokens	Defines the maximum number of tokens that the model can generate in its response. Setting a limit ensures that the response is concise and fits within the desired length constraints.
temperature	Controls the randomness of the model's output. A lower value close to 0 makes the output more deterministic, while a higher value close to 1 increases randomness and creativity. For example, if `temperature` is set to 0.5, the model balances between deterministic and creative outputs.
topP	Determines the cumulative probability threshold for token selection. The model considers the smallest set of tokens whose cumulative probability meets or exceeds `topP`. For example, if `topP` is set to 0.1, the model considers only the top 10% most probable tokens at each step.
topK	Limits the number of the highest-probability tokens to consider during response generation. For example, if `topK` is set to 2, the model considers only the top 2 tokens at each step, controlling output diversity and quality.

9. Save and publish the process.

Alternatively, you can also publish the **Search Tool Workflow** process from the assets page. To publish the process, click **Actions** in the row that contains the process and select **Publish**. The subprocesses will be published automatically. You can modify the process as mentioned in steps 7 and 8 before publishing the process if needed.

Invoking the process

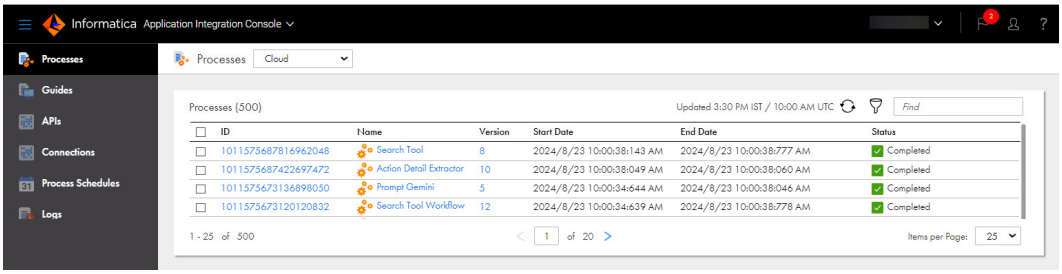
When you invoke the Simple React AI Agent with Search Tools process, the process orchestrates the whole workflow based on the LLM response.

1. Open the **Search Tool Workflow** process and click **Actions** > **Properties Details** > **Copy Service URL**.
2. Open a text editor and add the input fields and values to the service URL as shown in the following format:
   ```
   <Cloud Application Integration POD URL>/active-bpel/public/rt/<API_name>?
   Query=<User_Query>
   ```
3. Open a browser and paste the service URL.

You can also invoke the process using the Run Using option, REST or SOAP API endpoints in any API client, such as cURL, Postman, SOAP UI, or through any programming language.

In Application Integration Console, you can verify whether the process execution was successful or faulted.

The following image shows a successful process execution:


Informatica Application Integration Console
Processes
Guides
APIs
Connections
Process Schedules
Logs
Processes
Cloud
Processes (500)
Updated 3:30 PM IST / 10:00 AM UTC
Find
ID
Name
Version
Start Date
End Date
Status
1011575687816962048
Search Tool
8
2024/8/23 10:00:38:143 AM
2024/8/23 10:00:38:777 AM
Completed
1011575687422697472
Action Detail Extractor
10
2024/8/23 10:00:38:049 AM
2024/8/23 10:00:38:060 AM
Completed
1011575673136898050
Prompt Gemini
5
2024/8/23 10:00:34:644 AM
2024/8/23 10:00:38:046 AM
Completed
1011575673120120832
Search Tool Workflow
12
2024/8/23 10:00:34:639 AM
2024/8/23 10:00:38:778 AM
Completed
1 - 25 of 500
1
of 20
Items per Page:
25