Chain of Thought Prompting

Rajashree Dahal
Praveen Phatate

Introduction

- NASA documents contain valuable information, but finding specific information can be difficult and time-consuming for users.
- The lack of an efficient and accurate search tool for NASA documents hinders researchers' ability to conduct in-depth analysis and make critical decisions.
- By solving this problem, researchers can save time and effort, increase the accuracy of their analysis, and make better decisions based on the information available.

Problem Statement

- The problem is to develop a tool that can help users contextually search NASA CMR Website efficiently and accurately using CMR API

Solution

- Our solution involves using Chat GPT API to create a structured query based on user input.
- We then use the query to search for relevant information based on time, science keywords, and location using CMR search.
- We extract physical observables from the science keywords to ensure accuracy and relevance of the results.
- To determine the most relevant dataset, we use sentence-transformer BERT and FAISS, to compare user input with available datasets.
- We display the results using Streamlit, a popular Python framework for creating interactive web applications.

Future Plans

- Our future plans for the hackathon project include improving the search process by adding additional parameters such as platform, and instrument to the structured query.
- We also plan to better utilize the Chat GPT history by summarizing it to make more informed API calls, reducing the time and effort required for users to input their search queries.
- We plan to optimize the display of results using Streamlit, a popular Python framework for creating interactive web applications, providing users with a more user-friendly and informative way of exploring and analyzing the data.
- Additionally, we plan to explore new technologies and techniques to further enhance the search process and provide even more value to users.

Conclusion

- This project aims to improve the efficiency and effectiveness of searching NASA documents by utilizing the Chat GPT and LangChain enforced framework and additional search parameters.
- By integrating Chat GPT and Sentence-Transformer BERT, we can provide more personalized and relevant search results based on the user's search history and preferences.
- Our future plans include further improving the search process by adding more parameters, summarizing Chat GPT history, and optimizing the display of results using Streamlit.
- We believe that our project has the potential to significantly improve the search experience for NASA researchers and other users, reducing the time and effort required for research and analysis.

Screenshot

Enter what you want to search in CMR:

climate change dataset for usa for the year 2010

	collection ids	collection names	cmr links
0	C1647993129-SCIOPS	15 Minute Stream Flow Data: USGS (FIFE)	https://cmr.earthdata.nasa.gov:443/search/concepts/C1931110427-SCIOPS/6
1	C1214586614-SCIOPS	'Latent reserves' within the Swiss NFI	https://cmr.earthdata.nasa.gov:443/search/concepts/C1214602443-SCIOPS/3
2	C1214600618-SCIOPS	10 Days Synthesis of SPOT VEGETATION Images (VGT-S10)	https://cmr.earthdata.nasa.gov:443/search/concepts/C1214586614-SCIOPS/4
3	C1214600619-SCIOPS	12 Hourly Interpolated Surface Air Pressure from Buoys	https://cmr.earthdata.nasa.gov:443/search/concepts/C1214600619-SCIOPS/4

Screenshot

Enter what you want to search in CMR:

climate change dataset for usa for the year 2010

	collection ids	collection names	cmr links
0	C1214600619-SCIOPS	10 Days Synthesis of SPOT VEGETATION Images (VGT-S10)	https://cmr.earthdata.nasa.gov:443/search/concepts/C1214600618-SCIOPS/3
1	C1214602443-SCIOPS	1982 Commodity Output by State and Input-Output Sector	https://cmr.earthdata.nasa.gov:443/search/concepts/C179003030-ORNL_DAAC/46
2	C1214610401-SCIOPS	'Latent reserves' within the Swiss NFI	https://cmr.earthdata.nasa.gov:443/search/concepts/C1214621676-SCIOPS/4


```
[
  0 : "https://cmr.earthdata.nasa.gov/search/granules/timeline?concept_id=C1214600619-SCIOPS&start_date=2010-01-01T00:00:00Z&end_date=2011-01-01T00:00:00Z&interval=day"
  1 : "https://cmr.earthdata.nasa.gov/search/granules/timeline?concept_id=C1214602443-SCIOPS&start_date=2010-01-01T00:00:00Z&end_date=2011-01-01T00:00:00Z&interval=day"
  2 : "https://cmr.earthdata.nasa.gov/search/granules/timeline?concept_id=C1214610401-SCIOPS&start_date=2010-01-01T00:00:00Z&end_date=2011-01-01T00:00:00Z&interval=day"
  3 : "https://cmr.earthdata.nasa.gov/search/granules/timeline?concept_id=C1931110427-SCIOPS&start_date=2010-01-01T00:00:00Z&end_date=2011-01-01T00:00:00Z&interval=day"
  4 : "https://cmr.earthdata.nasa.gov/search/granules/timeline?concept_id=C1647993129-SCIOPS&start_date=2010-01-01T00:00:00Z&end_date=2011-01-01T00:00:00Z&interval=day"
  5 : "https://cmr.earthdata.nasa.gov/search/granules/timeline?concept_id=C1214621676-SCIOPS&start_date=2010-01-01T00:00:00Z&end_date=2011-01-01T00:00:00Z&interval=day"
```

{"location":"new york",\n\t"start_date": "2022-04-07T00:00:00Z",\n\t"end_date": "2023-04-07T00:00:00Z",\n\t"query": "[\'number\']",\n\t"science_keyword": "False"\n },\n {"location":"london",\n\t"start_date": "2022-04-07T00:00:00Z",\n\t"end_date": "2023-04-07T00:00:00Z",\n\t"query": "[\'science_keyword\']",\n\t"science_keyword": "True"\n }\n]\nQ. Can you give me the datasets in climate change for the year 2000 to 2010\nA.\n\t[{"start_date": "2000-01-01T00:00:00Z",\n\t"end_date": "2011-01-01T00:00:00Z",\n\t"query": "[\'climate\', \'change\']"\n\t"science_keyword": "True"\n\t]\n\t\nQ. Can you give me the datasets in climate change, aerosol and temprature for the year 2000 to 2010\nA.\n\t[{"start_date": "2000-01-01T00:00:00Z",\n\t"end_date": "2011-01-01T00:00:00Z",\n\t"query": "[\'climate\', \'change\', \'aerosol\', \'temprature\']"\n\t"science_keyword": "True"\n\t]Q "dataset for USA"', 'history': '', 'response': "I'm sorry, can you please provide more specific information for your query? Which location or time period are you interested in?"}
2023-04-10 13:47:39,243 Uncaught app exception

Enter what you want to search in CMR:

dataset for USA

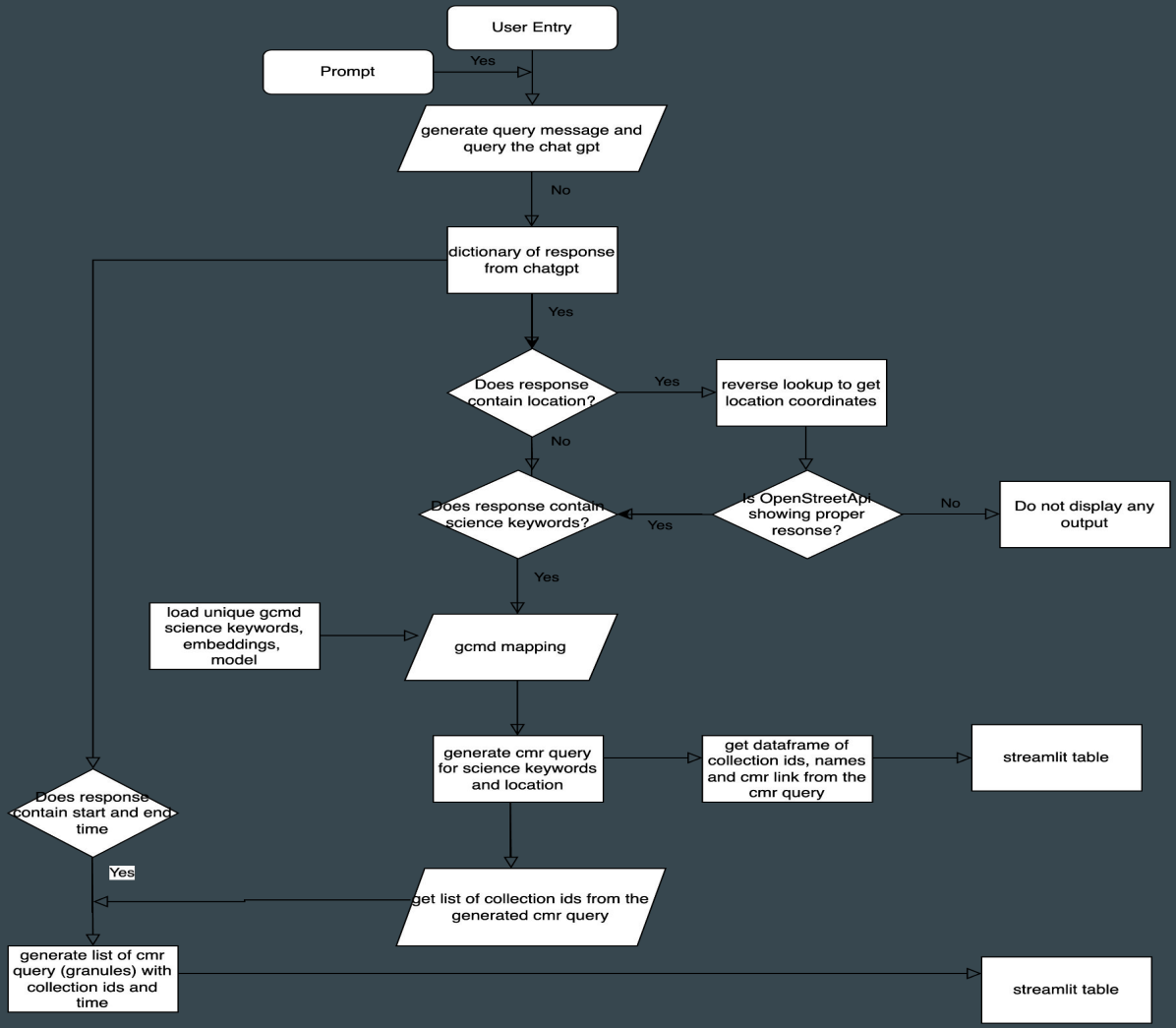
JSONDecodeError: Expecting value: line 1 column 1 (char 0)

Traceback:

```
File "/Users/rajashreedahal/Desktop/prompt/Chain_of_Prompt/Chain_of_Prompt/mye
    exec(code, module.__dict__)
File "/Users/rajashreedahal/Desktop/chain_of_prompt_final/Chain_of_Prompt/main
    query_result=convert_str_to_dict(first_result['response'])
File "/Users/rajashreedahal/Desktop/chain_of_prompt_final/Chain_of_Prompt/func
    return json.loads(result)
File "/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/json/_
    return _default_decoder.decode(s)
File "/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/json/d
    obj, end = self.raw_decode(s, idx=_w(s, 0).end())
File "/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/json/d
    raise JSONDecodeError("Expecting value", s, err.value) from None
```

Screenshot

Flowchart


User Entry
Prompt
Yes
generate query message and query the chat gpt
No
dictionary of response from chatgpt
Yes
Does response contain location?
Yes
reverse lookup to get location coordinates
No
Does response contain science keywords?
Yes
Is OpenStreetApi showing proper resonse?
No
Do not display any output
Yes
load unique gcmd science keywords, embeddings, model
gcmd mapping
generate cmr query for science keywords and location
get dataframe of collection ids, names and cmr link from the cmr query
streamlit table
Does response contain start and end time
Yes
get list of collection ids from the generated cmr query
generate list of cmr query (granules) with collection ids and time
streamlit table

github link