# Generating KGs from research papers using LLMs

# User Manual

Contents

[Generating KGs from research papers using LLMs 1](#_Toc212593808)

[User Manual 1](#_Toc212593809)

[1/ Introduction: 2](#_Toc212593810)

[a/ Overview of the system: 2](#_Toc212593811)

[b/ Purpose and Scope: 3](#_Toc212593812)

[c/ Intented Audience: 3](#_Toc212593813)

[2/ System requirements: 4](#_Toc212593814)

[a/ Hardware requirement: 4](#_Toc212593815)

[b/ Software requirement: 4](#_Toc212593816)

[3/ Getting started: 4](#_Toc212593817)

[a/ Installation: 4](#_Toc212593818)

[b/ Logging in: 4](#_Toc212593819)

[c/ Basic Navigation: 4](#_Toc212593820)

[4/ Using the system: 6](#_Toc212593821)

[a/ Extracting Cultural Knowledge Manually: 6](#_Toc212593822)

[b/ Generating Cypher queries using LLM: 6](#_Toc212593823)

[c/ Generating the KG using the Cypher queries: 7](#_Toc212593824)

[5/ Contact for Support: 7](#_Toc212593825)

## 1/ Introduction:

### a/ Overview of the system:

The system goal is to preserve cultural knowledge relating to fishing techniques that are gathered from fisherfolks and represent them in the most effective way possible. The system accomplishes this goal by extracting facts from research papers through the use of LLMs. The process is first done manually then it is automated by prompting the best performing model which is evaluated by the manual process.

The extracted facts are converted into relationships and cypher-like queries. Then, the Cypher queries are constructed by prompting the chosen LLM with the generated queries and relationships above. Finally, the Knowledge Graphs (KGs) are generated by inserting the generated queries into the Neo4j platform which is shown in the **Explores** option of Neo4j’s panel.

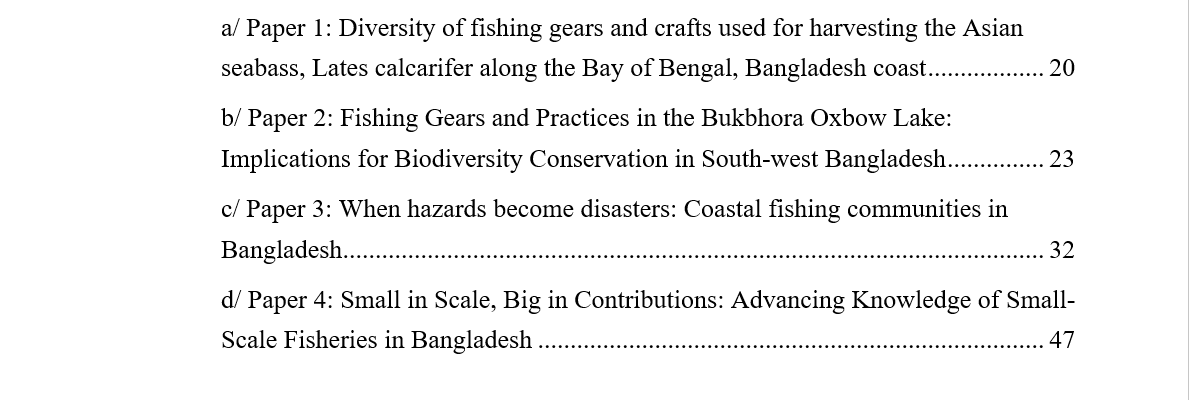


Figure 1 Overview of the manual facts extraction document

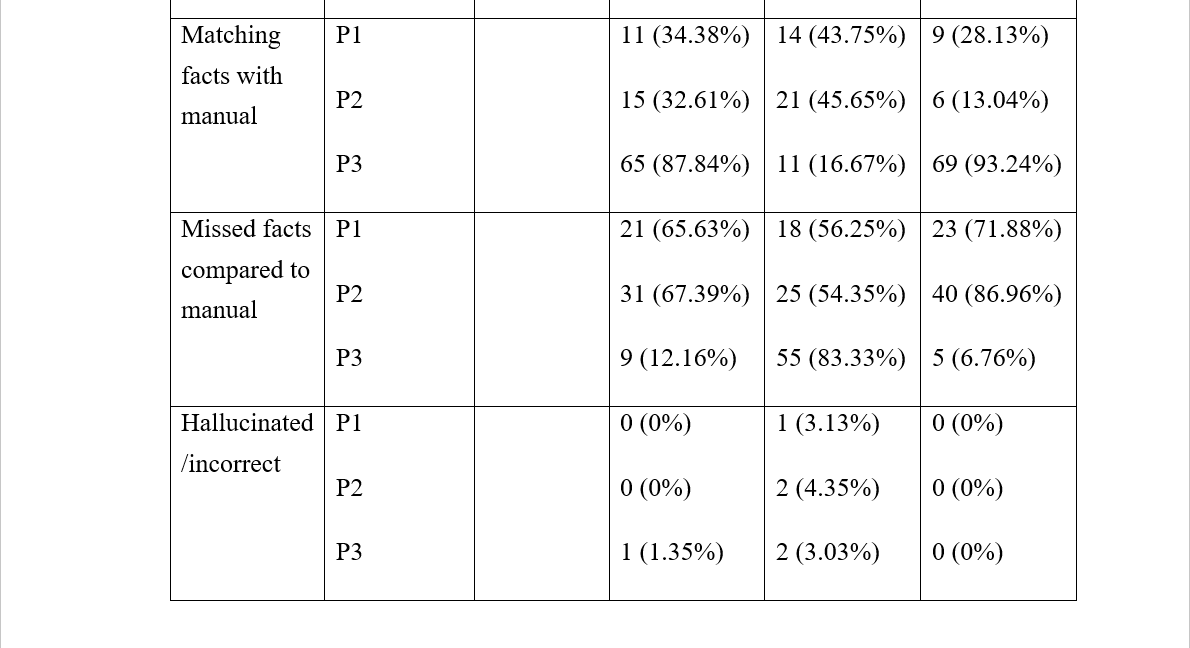
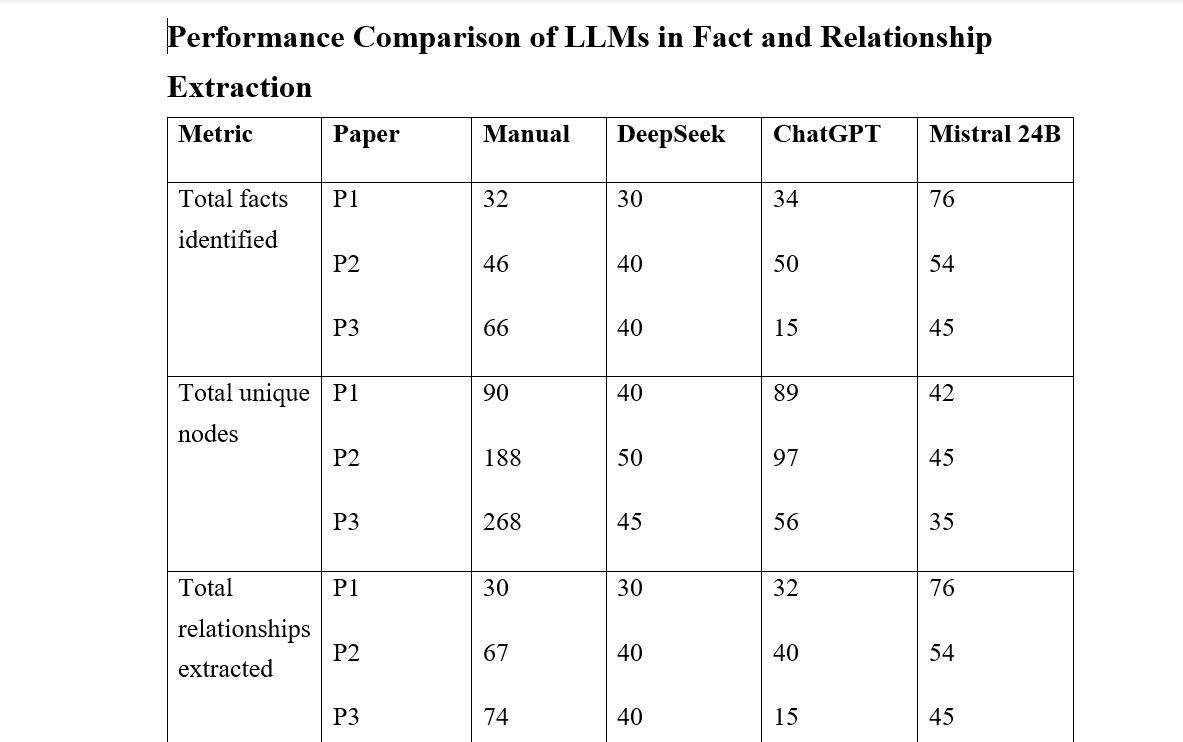


Figure 2 The comparison table for evaluating LLMs

### b/ Purpose and Scope:

This manual provides users with instructions on how to use a given research document to extract the relevant facts and relationships then use them to form Cypher queries which can be inserted into the Neo4j’s platform for KG creation. The instructions will be presented in an easy-to-understand way with minimal overview of the development process.

### c/ Intented Audience:

This manual is intended for any users of the system and should be kept nearby when operating the facts extraction and KG creation processes.

## 2/ System requirements:

### a/ Hardware requirement:

The required hardware for running the project: A working computer with internet access.

If the computer required internet access through wifi, the modem should be installed.

### b/ Software requirement:

The computer should have Windows 10 or above operating systems installed.

## 3/ Getting started:

### a/ Installation:

Follow these steps to download all key artifacts for operating the KG creation project:

* Step 1: Follow the link below to the Github repository for project B (in the search bar) - <https://github.com/SVA44/project-b-tacit-KG>
* Step 2: Download the documents *Manual Process of Extracting Facts and Relationships* and *Performance Comparison of LLMs in Fact and Relationship Extraction*
* Step 3: Scroll down to the README file and follow the listed links to all 3 chatbot models

### b/ Logging in:

The process is made publicly available via the Github repository: project-b-tacit-KG. Therefore, no logging in is required.

### c/ Basic Navigation:

In the document: *Manual Process of Extracting Facts and Relationships*, there are two tables for separating sentences and for extracting relationships from the facts sentences. For each document, there are 2 such corresponding tables. The document consists of tables for 4 research papers.

In the document: *Performance Comparison of LLMs in Fact and Relationship Extraction*, there is a table with statistics for the each chatbots present in the evaluation which are ChatGPT, Mistral 24B and DeepSeek. The benchmarks are evaluated for 3 papers which correspond to the first three papers from the manual process above.

The links to chatbot should contain an overview of a chat and a chat bar. An example of ChatGPT interface is given below:

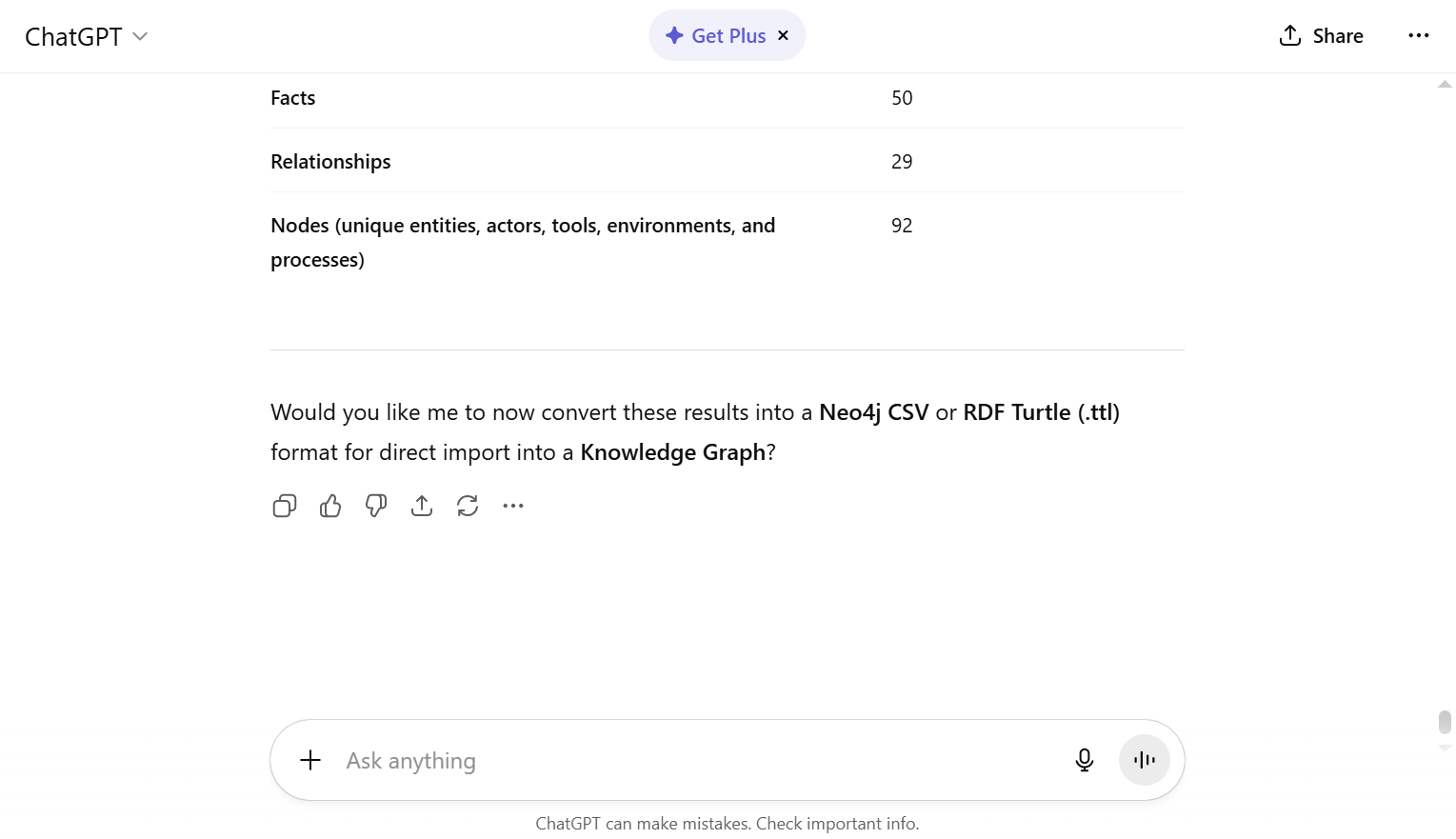


Figure 3 ChatGPT interface

The chat bar (at the bottom in the example) is used for typing prompts into chatbot. This will be used mainly for the extraction and queries formation.

The Neo4j platform is used for KG creation by copying the query generated above into the **Query** tab. The KG (Knowledge Graph) is shown in the **Explore** tab.

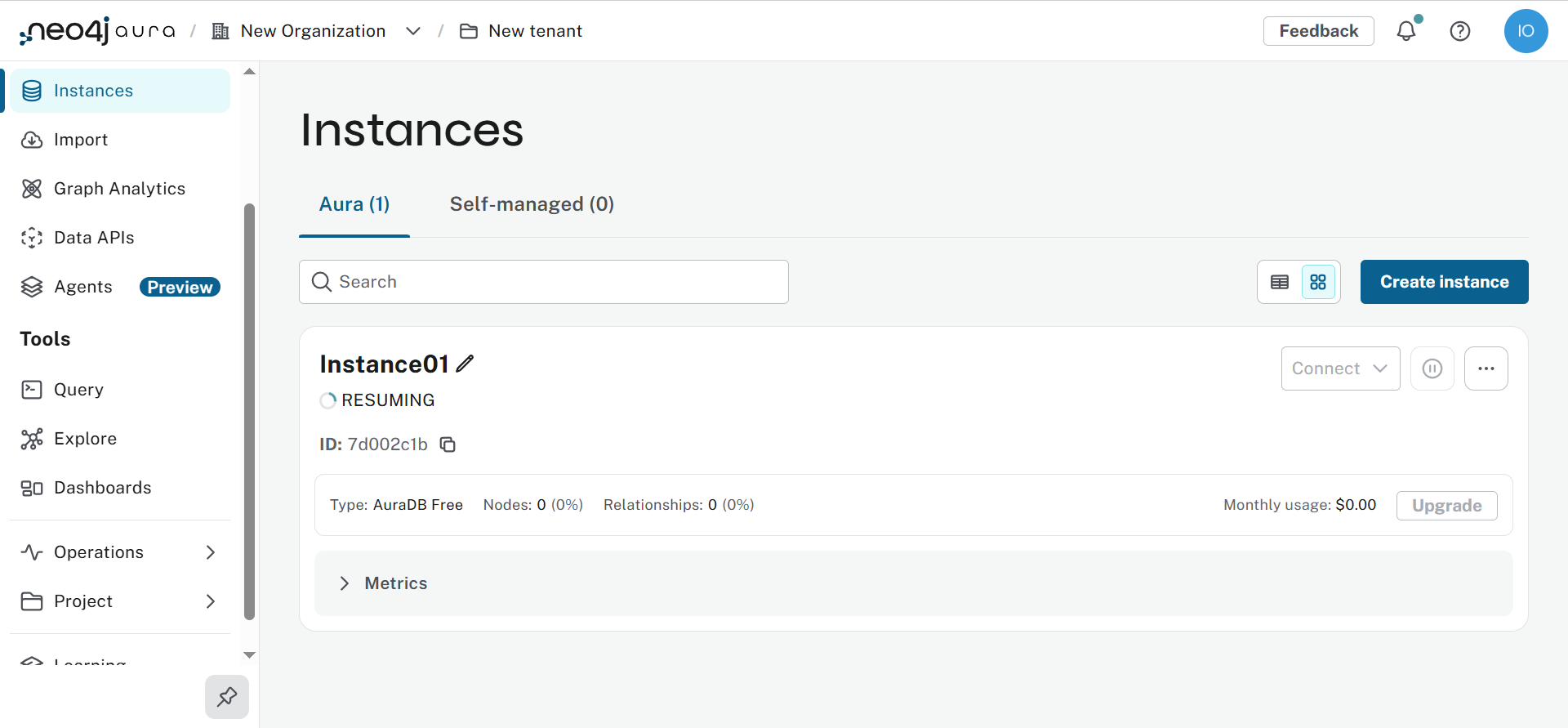


Figure 4 Neo4j platform interface

## 4/ Using the system:

### a/ Extracting Cultural Knowledge Manually:

Using the comparison table in the *Performance Comparison of LLMs in Fact and Relationship Extraction*, choose a model that is best suited for the facts’ extraction process. The recommendation is Mistral 24B which outperforms other models on its accuracy rates.

Follow the corresponding links to the chatbot’s interface then open the chat bar. Use the following prompt template:

**Prompt 1**: You are cultural knowledge research assistant. Your task is to extract all tacit fishing knowledge facts from the attached paper, then translate them into relationships. Your goal is to extract all important information for fisherfolk in [Country\_or\_region] to be stored in a Knowledge Graph (KG) and avoid all unrelated data. Present and number the facts extracted in a list. Finally, count the total number of nodes, relationships and facts. For example: [Example\_from\_the\_manual\_doc]. [Attached paper]

Fill in all the blank spaces with the appropriate information from the research paper then choose an example from the manual document. It is recommended to choose an example from the research paper of the manual process document that most closely matches the topic of the input research paper.

Click the arrow button to send the prompt to the model which should give an output.

### b/ Generating Cypher queries using LLM:

In the chat bar, enter the following prompt to generate the corresponding Cypher queries for the extracted facts above:

**Prompt 2**: You are cultural knowledge research assistant. Your goal is to extract all important information for fisherfolk in [Country\_or\_region] to be stored in a Knowledge Graph (KG) and avoid all unrelated data. Using the facts and relationships extracted above, generate the corresponding Cypher queries for the KG.

Again, fill in all the blank spaces with appropriate information from the research paper.

### c/ Generating the KG using the Cypher queries:

Copy all the Cypher queries code in the final response using the ‘Copy code’ button presented at the bottom of the codes’ space.

Then, open Neo4j console using the following link: <https://console-preview.neo4j.io/tools/query>. Open the **Query** tab and paste the content above into the query bar. Then press the play button at the right, this will generate the Knowledge Graph (KG):

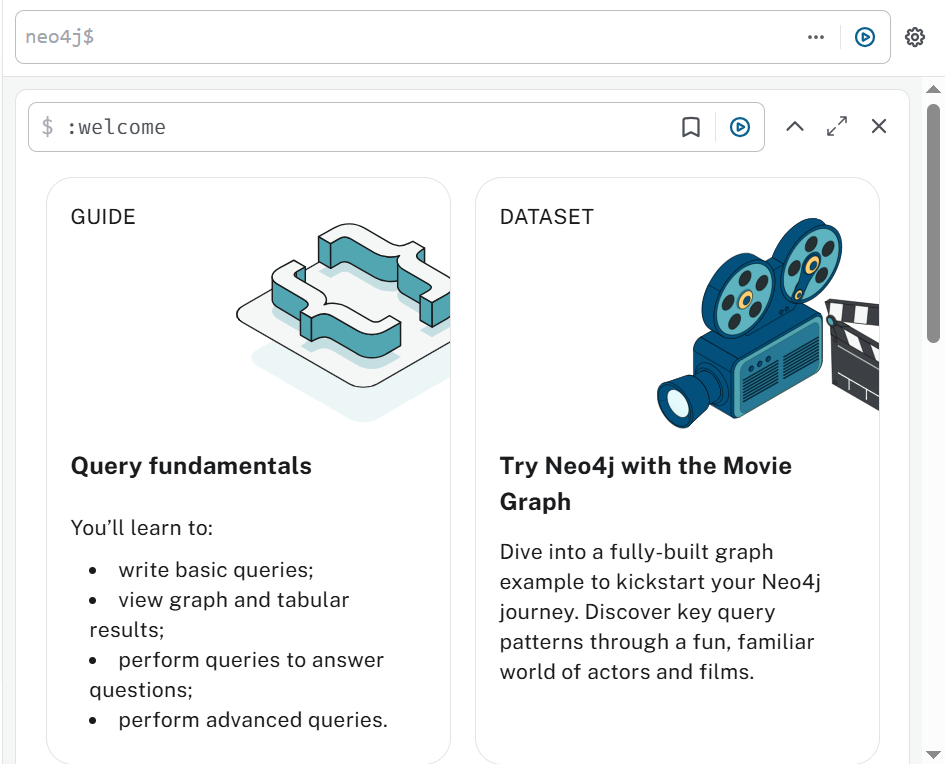


Figure 5 Neo4j query console

The resulting KG can be viewed by navigating to the **Explore** button below the **Query** button. An example is shown in the README file mentioned in the respository above.

## 5/ Contact for Support:

If there is any issues, contact developer at: [iom46905@gmail.com](mailto:iom46905@gmail.com).