Bridging the Gaps in LLM Interpretability of Unstructured Data with Knowledge Graphs

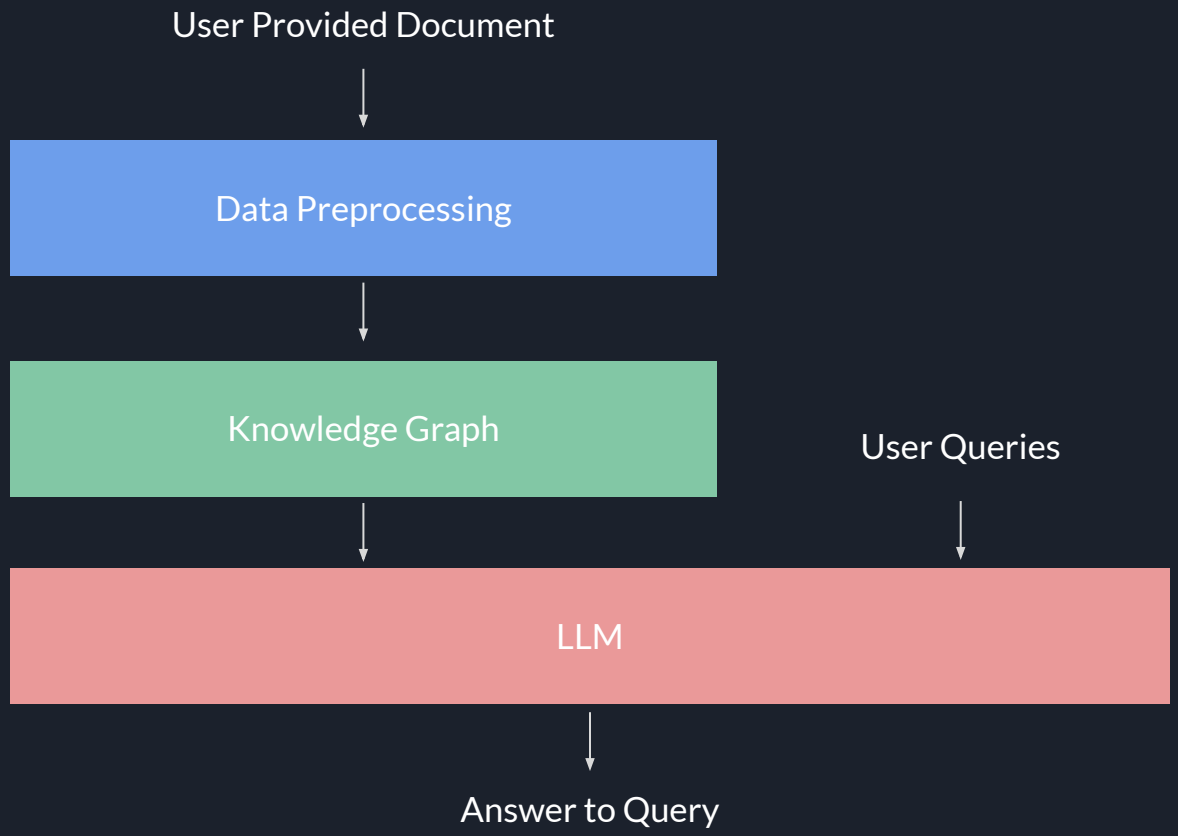
Eric Lin, Patrick Sun, Sean Wiryadi

Problem	Solution
Dependency on Extensive Data	Utilizing the context provided by the Knowledge Graph (KG)
Limited Understanding and Reasoning	Enhancing the LLM on a deeper context and clear data relationships
Lack of Generalizability	Use KG to improve LLM's adaptability and performance on things outside its training data

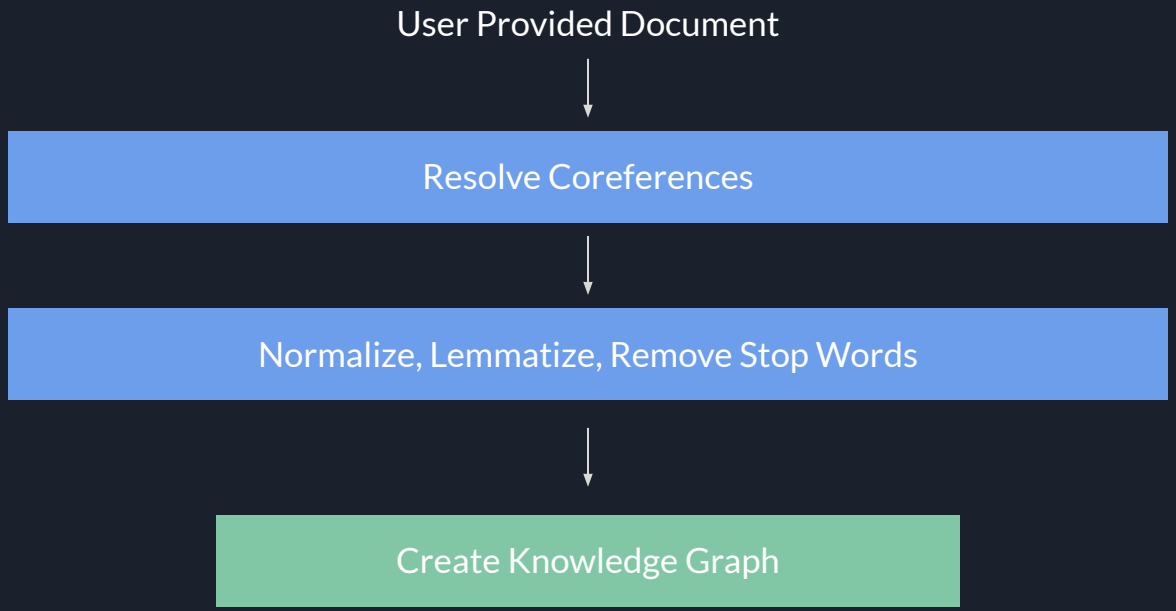
Outline

- Hybrid System
- Preprocessing methods
- Knowledge Graphs
- LLM Integration
- Evaluation Metrics

Hybrid System



Data Pre-Processing



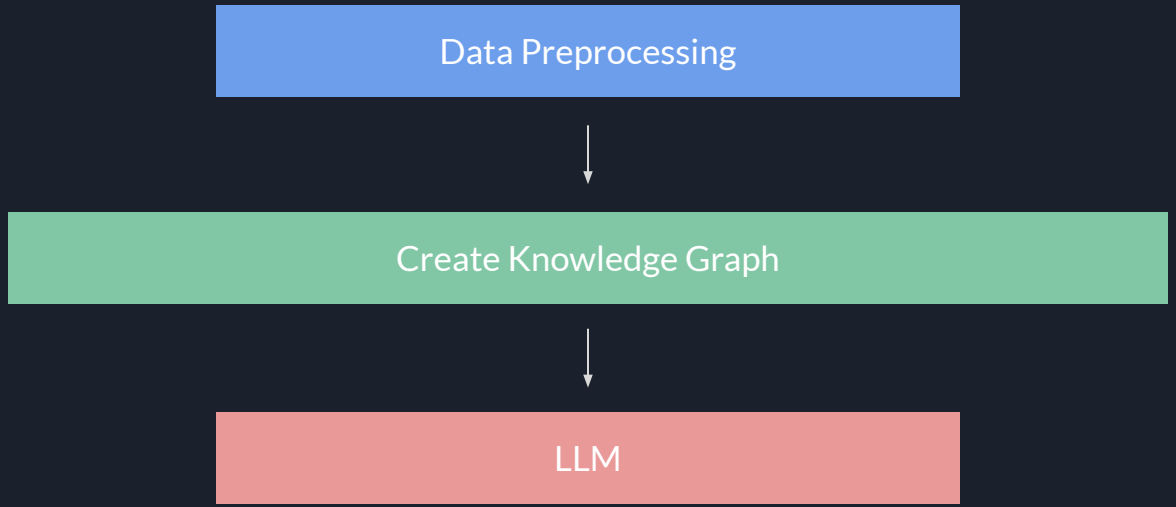
Resolving Coreferences

```
Example Input: I have a dog. He likes playing fetch.
Example Output: I have a dog. My dog likes playing fetch.

Example Input: John bought a new phone for Sandy. It was a gift for her.
Example Output: John bought a new phone for Sandy. The new phone was a gift for Sandy.

Example Input: Whenever I visit her, my grandma always gives me cookies. She is so sweet.
Example Output: Whenever I visit my grandma, my grandma always gives me cookies. My grandma is so sweet
```

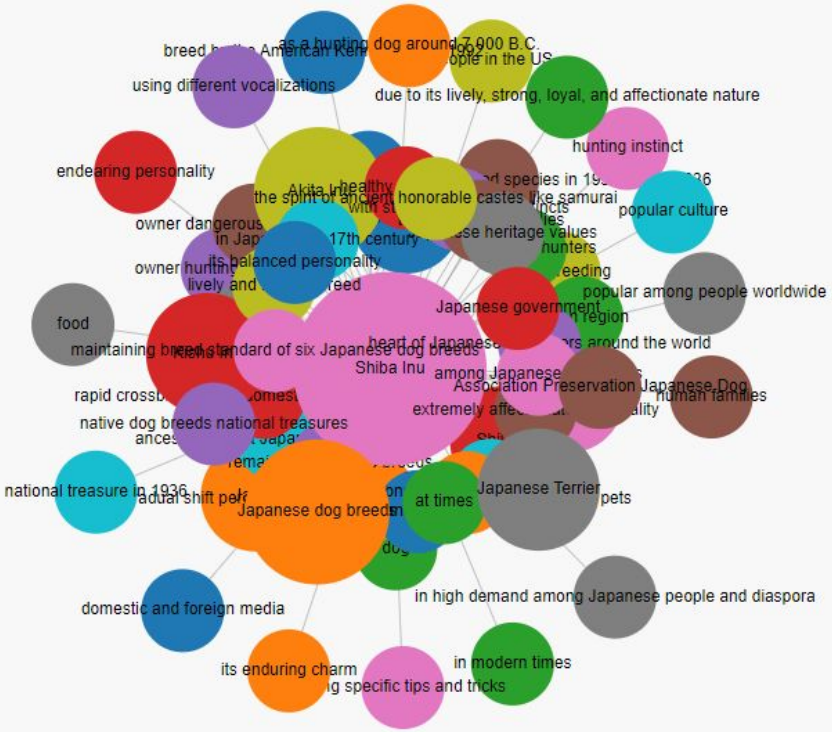
Creating the Knowledge Graph



(object1, relation, object2)

Example: (John, bought, house),
(house, built in, 2019)

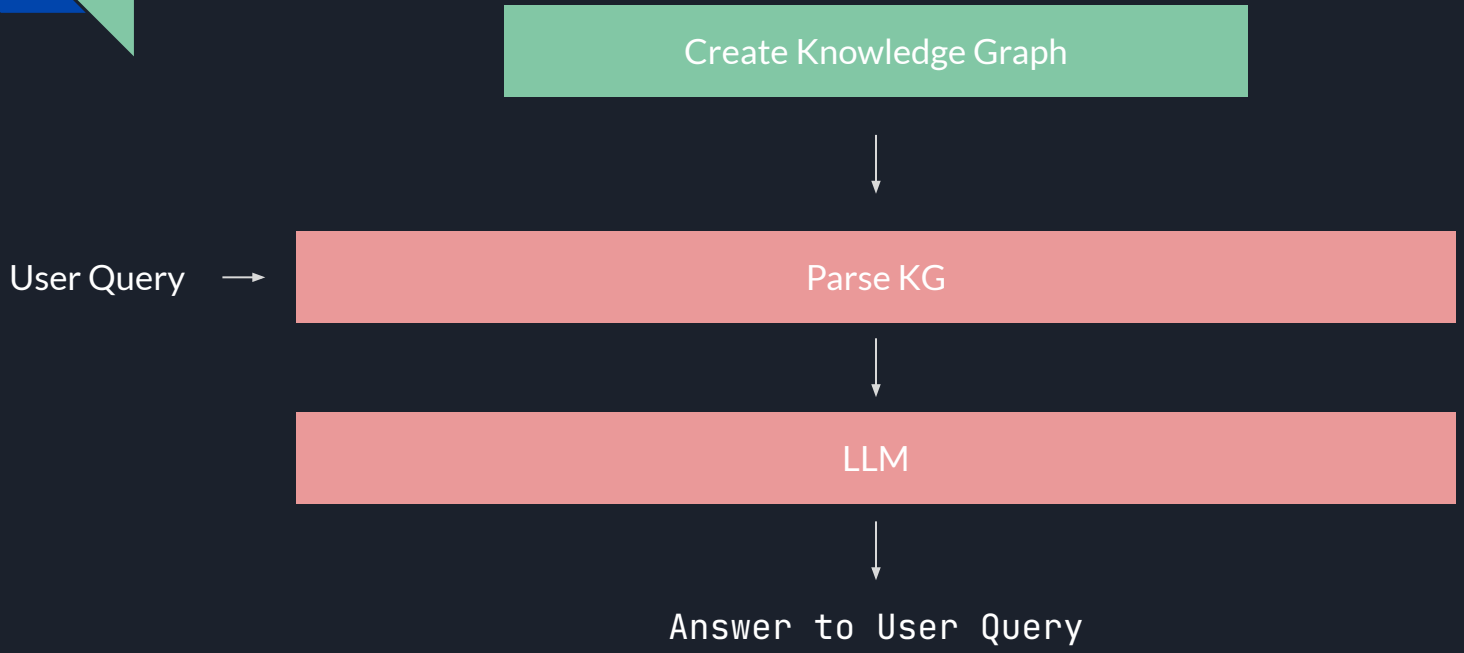
Set Context


using different vocalizations
due to its lively, strong, loyal, and affectionate nature
hunting instinct
endearing personality
popular culture
owner dangerous
its balanced personality
popular among people worldwide
food
Japanese government
Shiba Inu
native dog breeds national treasures
national treasure in 1936
Japanese dog breeds
Japanese Terrier
pets
at times
domestic and foreign media
in high demand among Japanese people and diaspora
its enduring charm
in modern times

Set Context

scientist
Niagara Falls
total eclipse
Helium
sun, moon, Earth align p
y afternoon
$6 billi
500,000 visitors
Companies
eclipse-theme
1868 eclipse
600 million years according to Johns Hopkins professor Sabine Stanley
30+ million people fro
America
US 2044
animal behavior as day suddenly turns to night
14 million visitors throughout e

LLM Integration



Tell me about shiba inus

Shiba Inus were declared national treasure in 1936, bred as hunting dogs around 7,000 B.C., known for their dominant and independent personality, loyalty to human families, love for food, and popular in global culture.

Tell me about akita inus

Akita Inus are revered for their large size, powerful frame, active and playful nature, vigilant personality, and possessiveness. They are often seen standing in a forest snow background.

Tell me about yorkshire terriers

No Info

Tell me about Japanese terriers

Japanese Terriers are recognized for their balanced personality, great hunting abilities, and being a lively and lovable breed.

When was the Association for the Preservation of the Japanese Dog founded

1928

Type your message here...

Send

How many people will see the eclipse

30+ million people

How much money will the eclipse bring to the economy?

$6 billion infusion US economy

What happened during the 1868 eclipse

helium was discovered

Type your message here...

Send

Evaluation

question string · *lengths* 22 4.24k	answer string · *lengths* 2 5.09k	date string · *lengths* 0 30	transcript string · *lengths* 1 4.1k	q string · *classes* 19 values	ticker string · *lengths* 1 6	predictions list · *lengths* 2 2
What was TSMC's revenue in US dolla…	I do not know. The transcript does not…	Jan 13, 2022, 1:00 a.m. ET	and for our industry-leading…	2021-Q4	TSM	[{ "class_id": 0, "class_name":…
What was TSMC's EPS in 2019	I do not know. The transcript does not…	Jan 13, 2022, 1:00 a.m. ET	and for our industry-leading…	2021-Q4	TSM	[{ "class_id": 0, "class_name":…
What was TSMC's capex spending in…	I do not know. The transcript does not…	Jan 13, 2022, 1:00 a.m. ET	and for our industry-leading…	2021-Q4	TSM	[{ "class_id": 0, "class_name":…
What is the expected growth rate of…	The expected growth rate of global…	Jan 14, 2021, 1:00 a.m. ET	g demand for our advanced…	2020-Q4	TSM	[{ "class_id": 0, "class_name":…
What is the expected penetration rate fo…	The expected penetration rate fo…	Jan 14, 2021, 1:00 a.m. ET	g demand for our advanced…	2020-Q4	TSM	[{ "class_id": 0, "class_name":…
What is the expected revenue growth rate…	The expected revenue growth rate for TSM…	Jan 14, 2021, 1:00 a.m. ET	g demand for our advanced…	2020-Q4	TSM	[{ "class_id": 0, "class_name":…
What is TSMC's expected revenue…	The expected revenue growth rate for TSM…	Jul 14, 2022, 2:00 a.m. ET	ferentiating factors, we expect…	2022-Q2	TSM	[{ "class_id": 0, "class_name":…
What is the expected delivery schedule…	The expected delivery schedule…	Jul 14, 2022, 2:00 a.m. ET	ferentiating factors, we expect…	2022-Q2	TSM	[{ "class_id": 0, "class_name":…
What is the expected revenue contributio…	The expected revenue contribution from N…	Jul 14, 2022, 2:00 a.m. ET	ferentiating factors, we expect…	2022-Q2	TSM	[{ "class_id": 0, "class_name":…