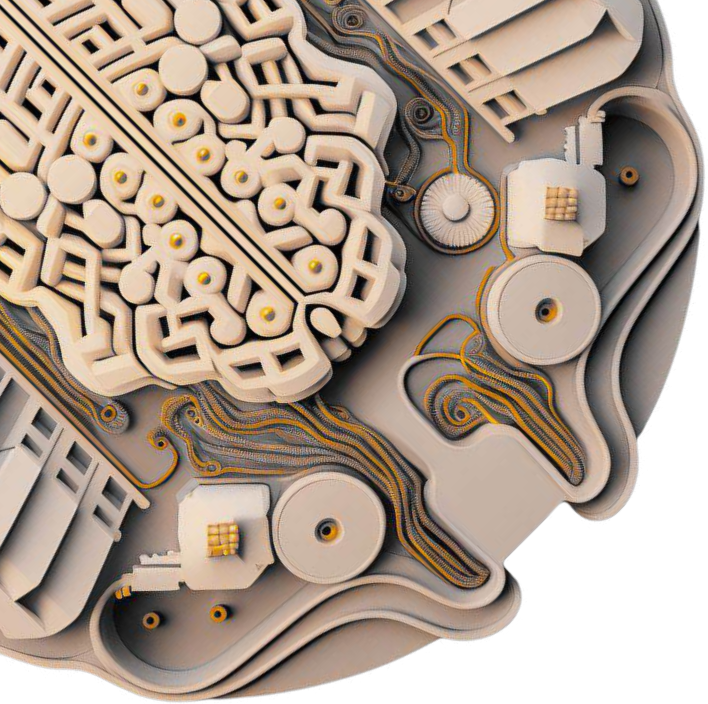


DATA INNOVATION SUMMIT 2024 MEA

EQT

MOTHERBRAIN

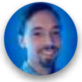
Knowledge Graph in Private Equity: LLM-enhanced Construction, Contextual Retrieval, and Reasoning

Lele Cao

Ph.D., Principal AI Research Scientist

lele.cao@eqtpartners.com

Credits also to:



Mark Granroth-Wilding

Valentin Buchner

Vilhelm von Ehrenheim

Drew McCornack

EQT & Motherbrain

- EQT: Private Equity Fund
  - Global investment fund w €200Bn+ AUM
  - Buys, grows & sells companies
  - Venture Capital, Buyout, Infrastructure, etc.

- Motherbrain: Data & Machine Learning Platform
  - Support investment professionals globally
  - Merging and enriching data using AI (on GCP)

- Read more about us → https://motherbrain.ai

Agenda

Motivation

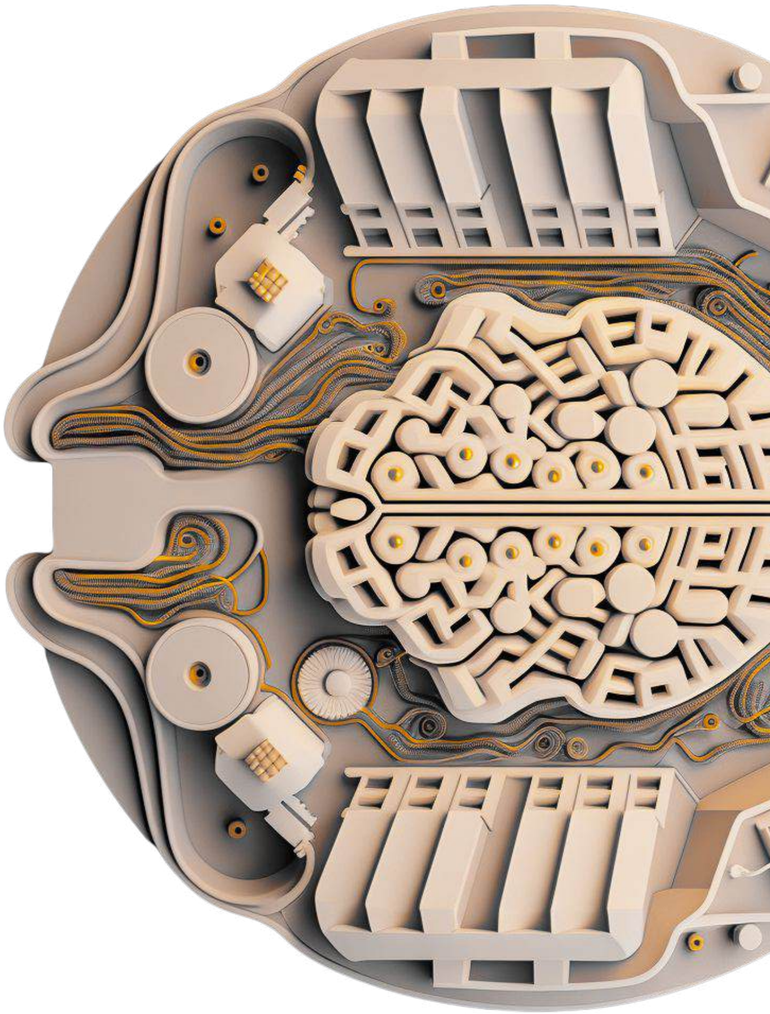
Approach

KG Construction

Contextual Retrieval

Reasoning

Summary



Motivation: a simplified view of deal process

Sourcing

- Which companies exists within my mandate?
- How do I make sure to spend time on interesting opportunities?
- What does the market look like around a deal?

Due Diligence

- What are the risks and attractions in this opportunity?
- What is the full potential this company can achieve?
- Who do we know we can ask for insights?

Holding

- How can we make sure to fulfil the full potential plan?
- What are the potential disruptions and market trends?
- How are we actually tracking on the plan?

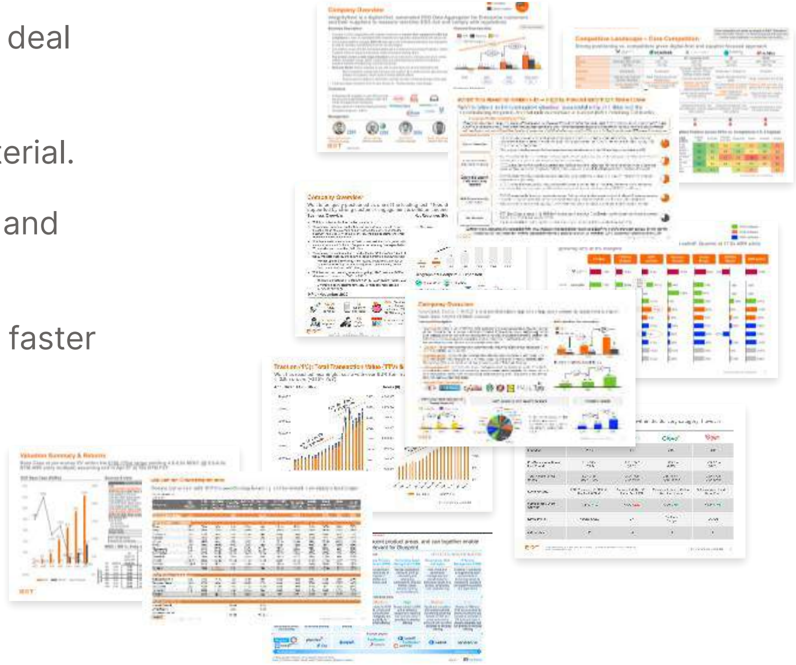
Exit

- How do we position the deal up for a successful sale?
- How can we make sure the company is in good health and can continue to grow?
- Which bankers and advisors should I talk to?

Motivation: leverage proprietary knowledge

Proprietary Knowledge

- So much interesting information is gathered throughout a deal process.
- EQT has been doing this since 1994… there is a lot of material.
- Reading through material is pretty intense as most decks and market reports are very dense. And long…
- But understanding historical deals is very important to be faster and smarter in future deal assessments.



Motivation: leverage third-party knowledge

Third-party Knowledge

- External experts and consultants often provide valuable insights that are not available internally.
- Leverage industry specialists can bring in-depth market analyses and fresh perspectives.
- Third-party data sources can offer objective viewpoints, free from internal biases.
- Most subscription-based sources are structured, hence easier to integrate.

PitchBook crunchbase ... ...

Agenda

Motivation

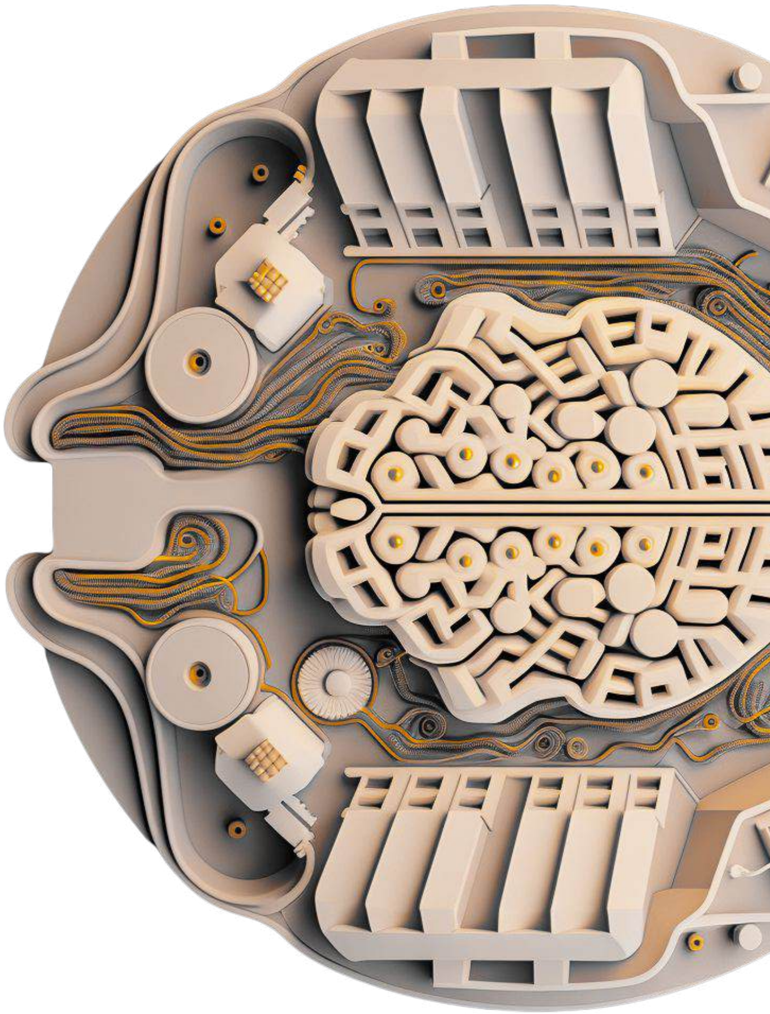
Approach

KG Construction

Contextual Retrieval

Reasoning

Summary



Approach: objective and applications

Develop an Intelligent and Knowledgeable Agent/Service to help deal professionals in many applications:

Ask a question about

- Mergers & Acquisitions
- Competitor Analysis
- Market Mapping
- Due Diligence Support
- Portfolio Company Growth Strategy
- Investment Thesis Formulation
- ... ...

→

→

Answer
here are some suggestions for deal advisers ... ...

Approach: a concrete example

Develop an Intelligent and Knowledgeable Agent/Service to help deal professionals in many applications:

Ask a question
We are looking at a deal of a company named ManyPets (manypets.com). Can you suggest some deal experts in related sectors and also some direct competitors in EU?

→

→

Answer
here are some suggestions for deal advisers in the pet insurance sector, as well as some direct competitors in the EU:
...

Approach: LLM prompting

Ask a question
We are looking at a deal of a company named ManyPets (manypets.com). Can you suggest some deal experts in related sectors and also some direct competitors in EU?

LLMs
ChatGPT
LLaMA
Gemini
Mistral
Claude
...

Answer
here are some suggestions for deal advisers in the pet insurance sector, as well as some direct competitors in the EU:
...

Approach: LLM prompting

Ask a question
We are looking at a deal of a company named ManyPets (manypets.com). Can you suggest some deal experts in related sectors and also some direct competitors in EU?

LLMs
ChatGPT
LLaMA
Gemini
Mistral
Claude
...

Answer
here are some suggestions for deal advisers in the pet insurance sector, as well as some direct competitors in the EU:
...

Problem: Lack domain-specific and up-to-date knowledge.

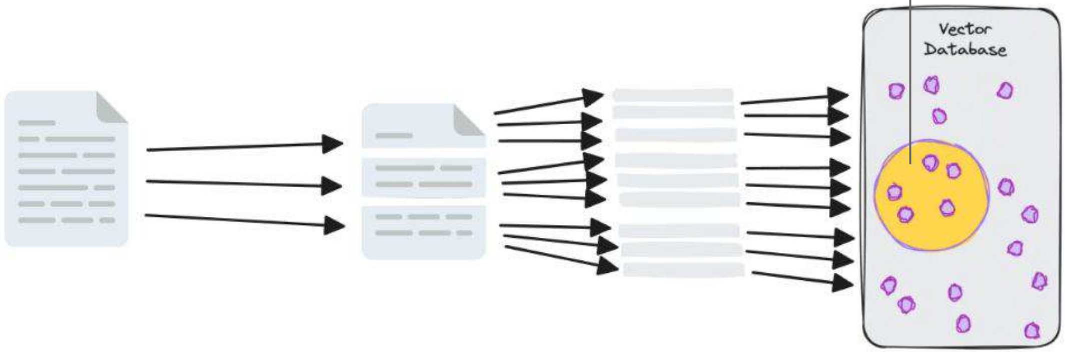
Approach: RAG - retrieval augmented generation

Ask a question
We are looking at a deal of a company named ManyPets (manypets.com). Can you suggest some deal experts in related sectors and also some direct competitors in EU?

LLMs
ChatGPT
LLaMA
Gemini
Mistral
Claude
...

Answer
here are some suggestions for deal advisers in the pet insurance sector, as well as some direct competitors in the EU:
...

Retrieve document slices



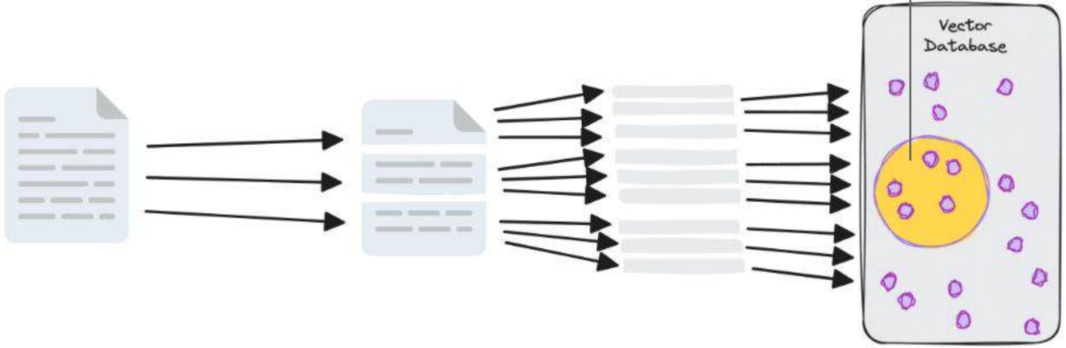
Approach: RAG - retrieval augmented generation

Ask a question
We are looking at a deal of a company named ManyPets (manypets.com). Can you suggest some deal experts in related sectors and also some direct competitors in EU?

LLMs
ChatGPT
LLaMA
Gemini
Mistral
Claude
...

Answer
here are some suggestions for deal advisers in the pet insurance sector, as well as some direct competitors in the EU:
...

Retrieve document slices



Problem:
- not modeling for PE-specific entities and their relationships;
- hard to integrate the third-party knowledge.

Approach: RAG over KG (knowledge graph)

Ask a question
We are looking at a deal of a company named ManyPets (manypets.com). Can you suggest some deal experts in related sectors and also some direct competitors in EU?

LLMs
ChatGPT
LLaMA
Gemini
Mistral
Claude
...

Reasoning

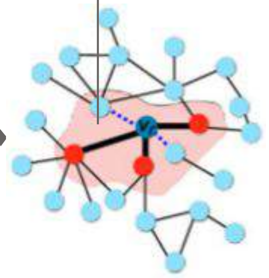
Answer
here are some suggestions for deal advisers in the pet insurance sector, as well as some direct competitors in the EU:
...

Retrieve relevant sub-graph

Third-Party Knowledge (Structured Data)

Construct

Proprietary Knowledge (Unstructured Docs)



Knowledge Graph

Approach: RAG over KG (knowledge graph)

Three Key Components:

1. **KG Construction** - Extract relevant entities, relations, and attributes from **proprietary documents** and **third-party data**.
2. **Contextual Retrieval** - According to the context provided by the query/question, retrieve the relevant sub-KG for LLM to reason about.
3. **Reasoning** - With the query/question and the retrieved sub-KG, generate the response/answer.

Agenda

Motivation

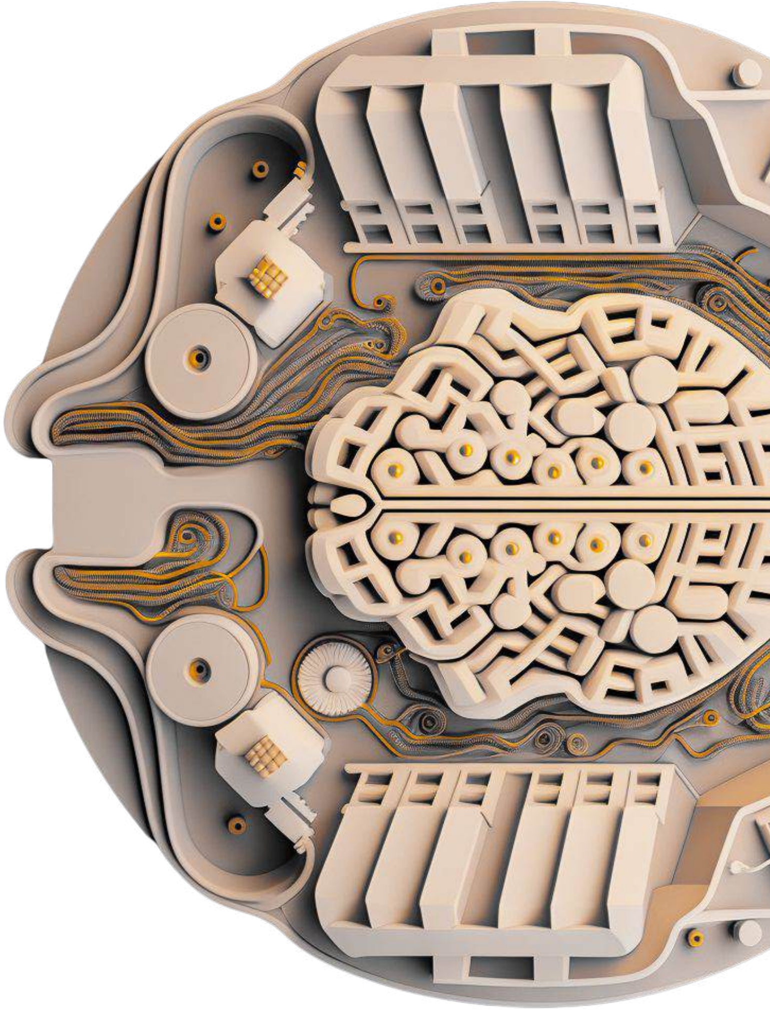
Approach

KG Construction

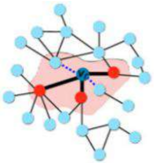
Contextual Retrieval

Reasoning

Summary

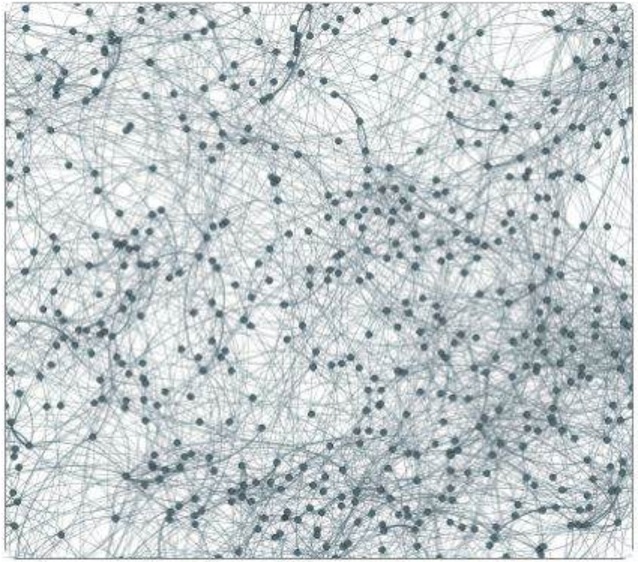


Third-Party Knowledge (Structured Data) $\xrightarrow{\text{Construct}}$



KG Construction: CompanyKG

- **Purpose**: quantify similarities among companies.
- **Example use cases**:
    - Market mapping
    - Competitor analysis
    - Mergers and Acquisitions (M&A)
- **Data source**: mostly third-party structure data, such as Pitchbook, Crunchbase, etc.

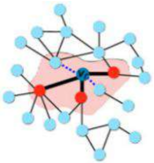


A fraction of the entire graph

https://github.com/EQTPartners/CompanyKG

Third-Party Knowledge (Structured Data)

Construct

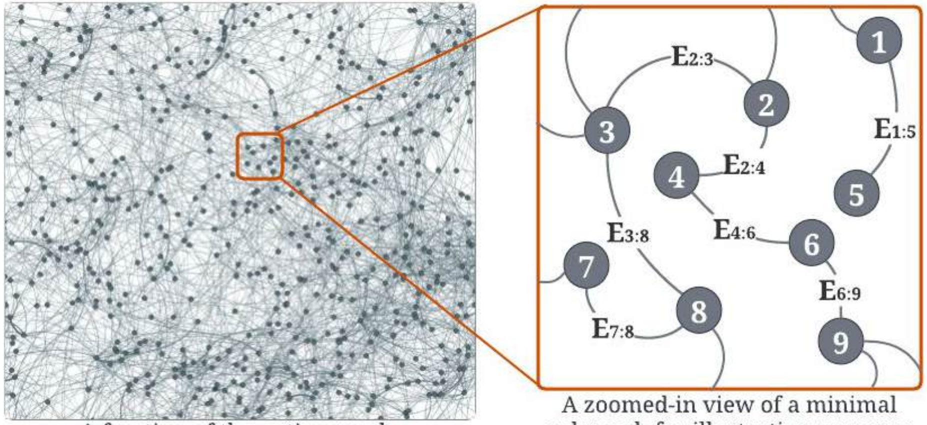


KG Construction: CompanyKG

Relations: 15 relation types in 6 categories

- competitive landscape
- industry sector
- M&A transactions
- people's affiliation
- news/event engagement
- and product positioning

51.06 million weighted edges



A fraction of the entire graph

A zoomed-in view of a minimal subgraph for illustrative purposes

Edge weights: 15-dim vector, where the *i*-th dimension is the weight of the *i*-th edge type (ET*i*)

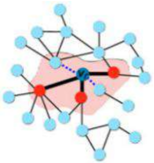
$E_{1:5}$ [0.0, 1.0, 0.0, 2.0, ... 5.1, 0.0, 1.6] $\in \mathbb{R}^{15}$

$E_{2:3}$ [0.0, 1.0, 1.0, 0.0, ... 0.0, 0.0, 2.3] $\in \mathbb{R}^{15}$

⋮

Third-Party Knowledge (Structured Data)

Construct

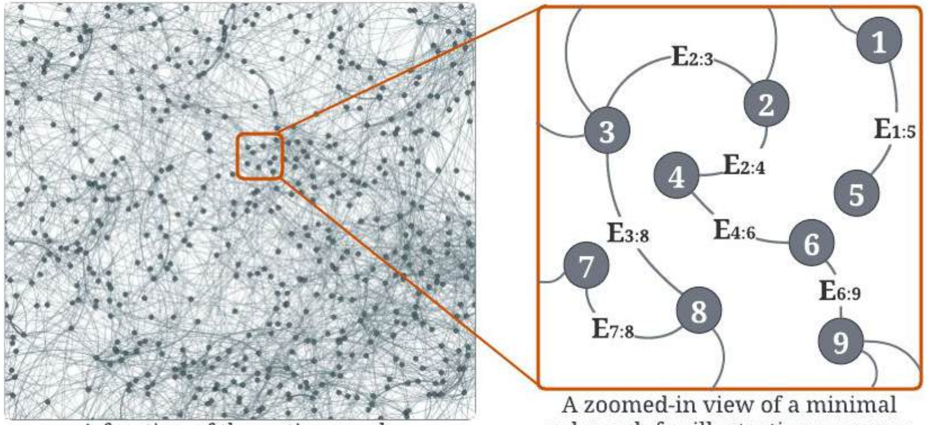


KG Construction: CompanyKG

Nodes: 1.17 million companies

Node feature: description/keywords embeddings:

- multilingual BERT
- ADA2 (GPT3.5)
- SimCSE
  https://doi.org/10.18653/v1/2021.emnlp-main.552
- PAUSE
  https://doi.org/10.18653/v1/2021.emnlp-main.791



A fraction of the entire graph

A zoomed-in view of a minimal subgraph for illustrative purposes

Node features: company description embeddings from 4 language models

1
mSBERT: $[0.32, 0.01, \ldots -0.49] \in \mathbb{R}^{512}$
ADA2: $[0.05, 0.20, \ldots\ 0.35] \in \mathbb{R}^{1536}$
SimCSE: $[0.29, 0.16, \ldots -0.24] \in \mathbb{R}^{768}$
PAUSE: $[0.02, 0.73, \ldots\ 0.88] \in \mathbb{R}^{32}$

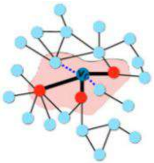
2
mSBERT: $[0.90, 0.53, \ldots\ 0.05] \in \mathbb{R}^{512}$
ADA2: $[0.44, -0.10, \ldots\ 0.35] \in \mathbb{R}^{1536}$
SimCSE: $[0.83, 0.01, \ldots\ 0.54] \in \mathbb{R}^{768}$
PAUSE: $[-0.22, 0.06, \ldots\ 0.90] \in \mathbb{R}^{32}$

⋮

https://github.com/EQTPartners/CompanyKG

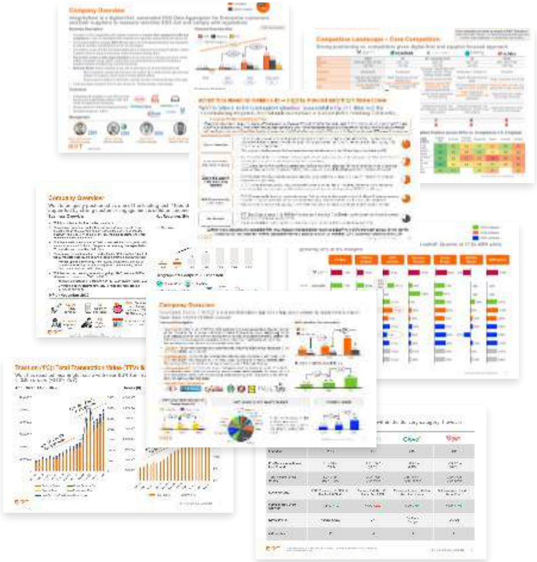
Proprietary Knowledge
(Unstructured Docs)

Construct



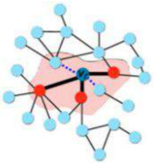
KG Construction: PEKG

- **Purpose**: Construct Knowledge Graph for EQT's PE deals.
- **Data source**: EQT's proprietary deal related documents. Each document is about a specific company (a.k.a., target company) in scope.



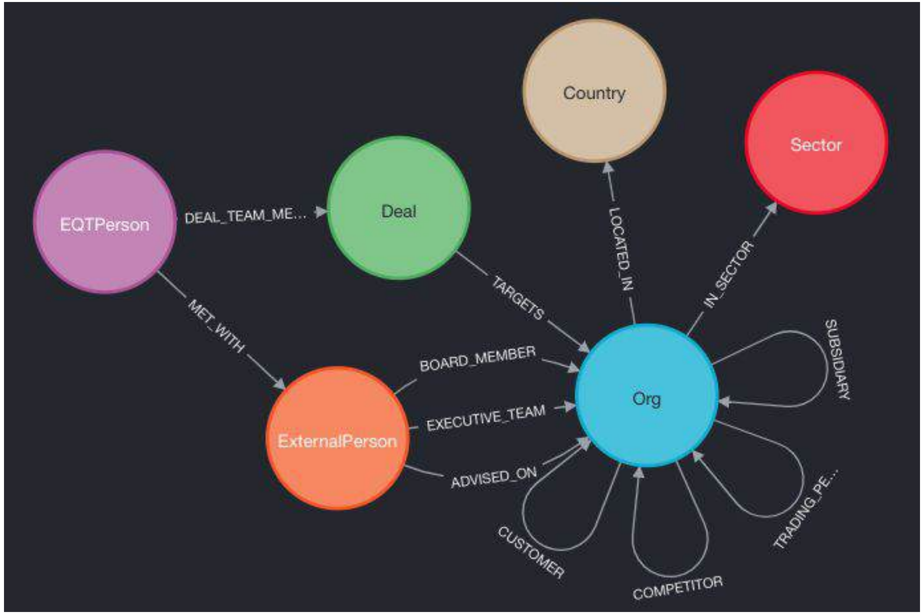
EQT's PE
deal docs

Proprietary Knowledge (Unstructured Docs) **Construct**

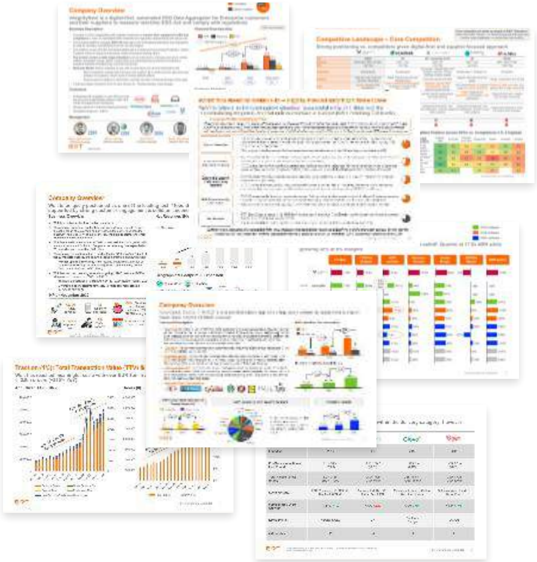


KG Construction: PEKG

- What **entities** and **relations** we extract and build into PEKG?



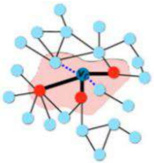
Extraction Guide (Meta Graph)



EQT's PE deal docs

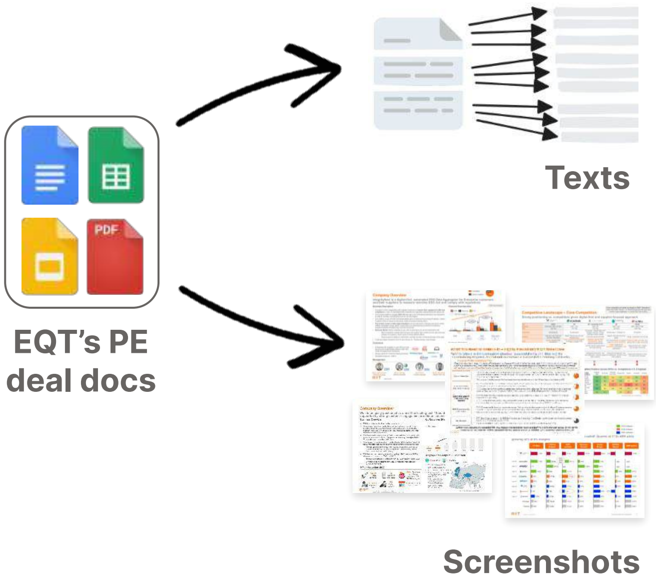
Proprietary Knowledge
(Unstructured Docs)

Construct



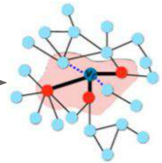
KG Construction: PEKG

- How do we automate PEKG construction?



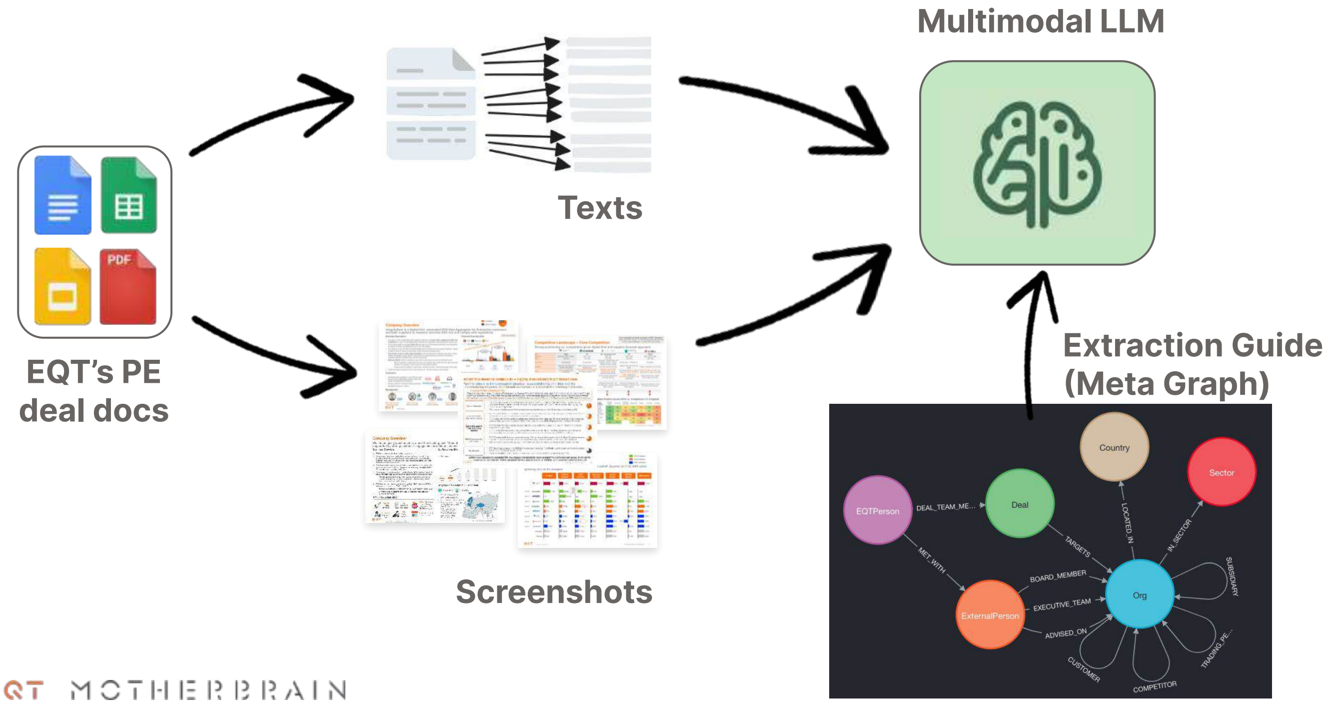
Proprietary Knowledge
(Unstructured Docs)

Construct



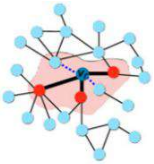
KG Construction: PEKG

- How do we automate PEKG construction?



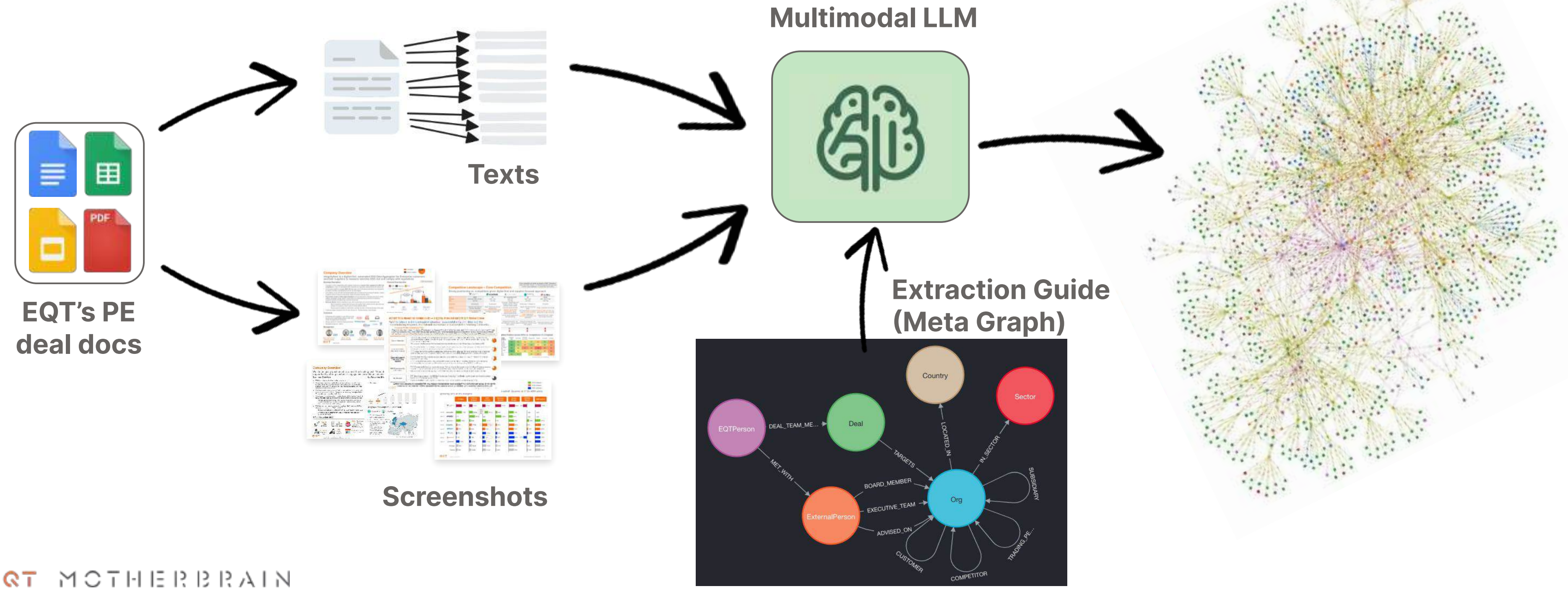
Proprietary Knowledge
(Unstructured Docs)

Construct

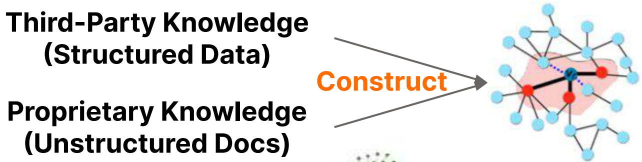


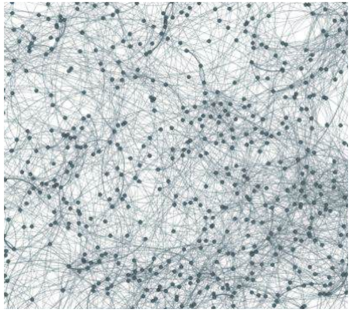
KG Construction: PEKG

- How do we automate PEKG construction?

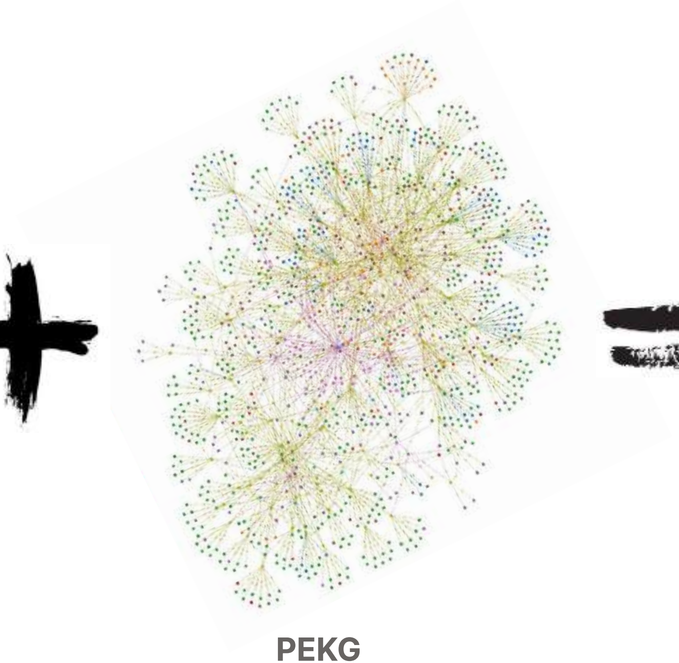


KG Construction: PEKG + CompanyKG

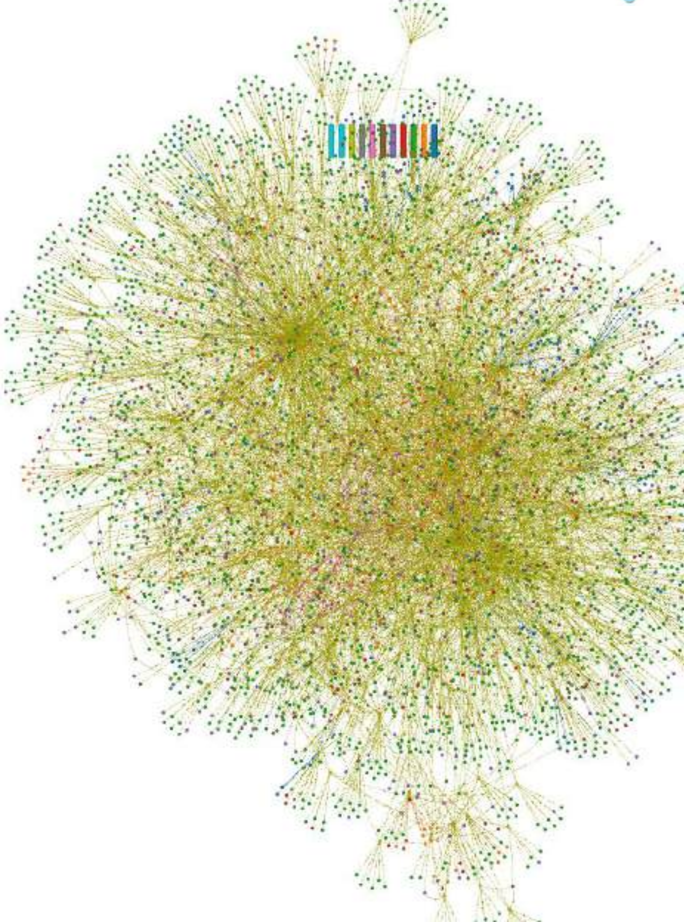




CompanyKG



PEKG



Agenda

Motivation

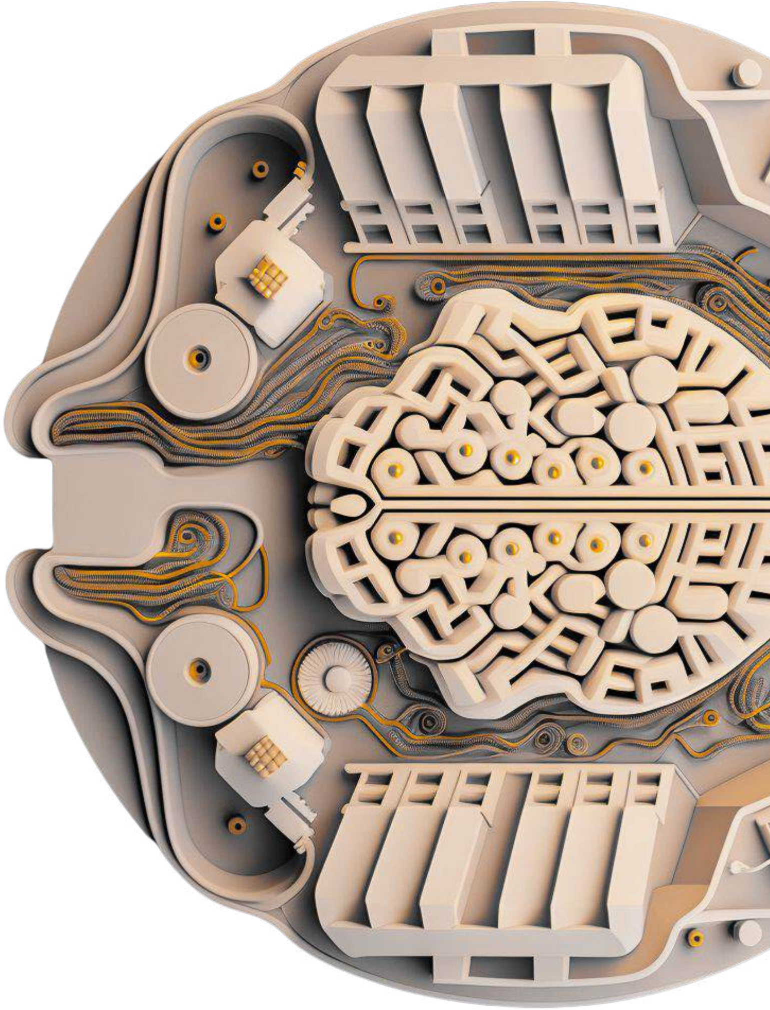
Approach

KG Construction

Contextual Retrieval

Reasoning

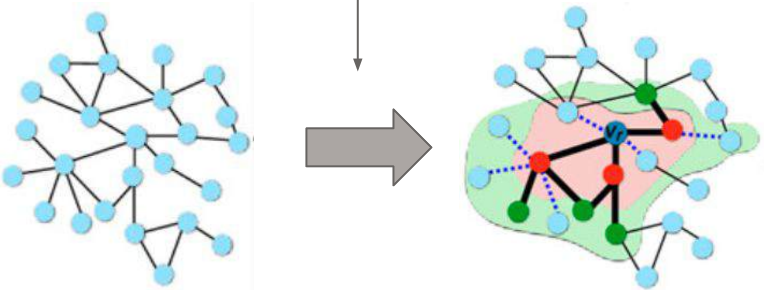
Summary



Contextual Retrieval

Objective: retrieve a sub-KG from the entire KG according to the context of the input query/question.

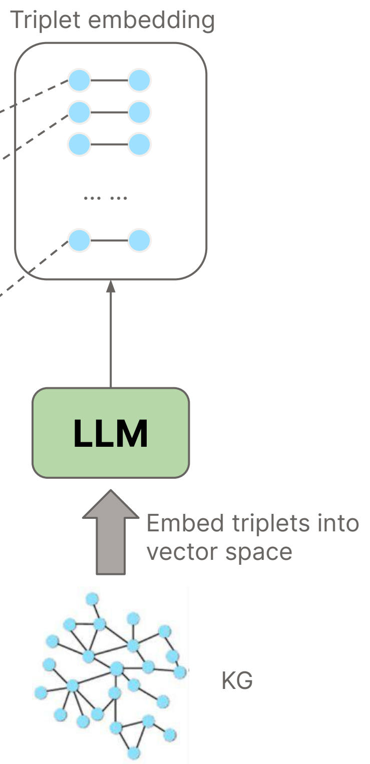
- The existing work mostly
  - assumes the availability of a contextual sub-KG, which is not true in reality;
  - or adopt a overly simplified approach, such as randomly expand 2 steps from a center node.



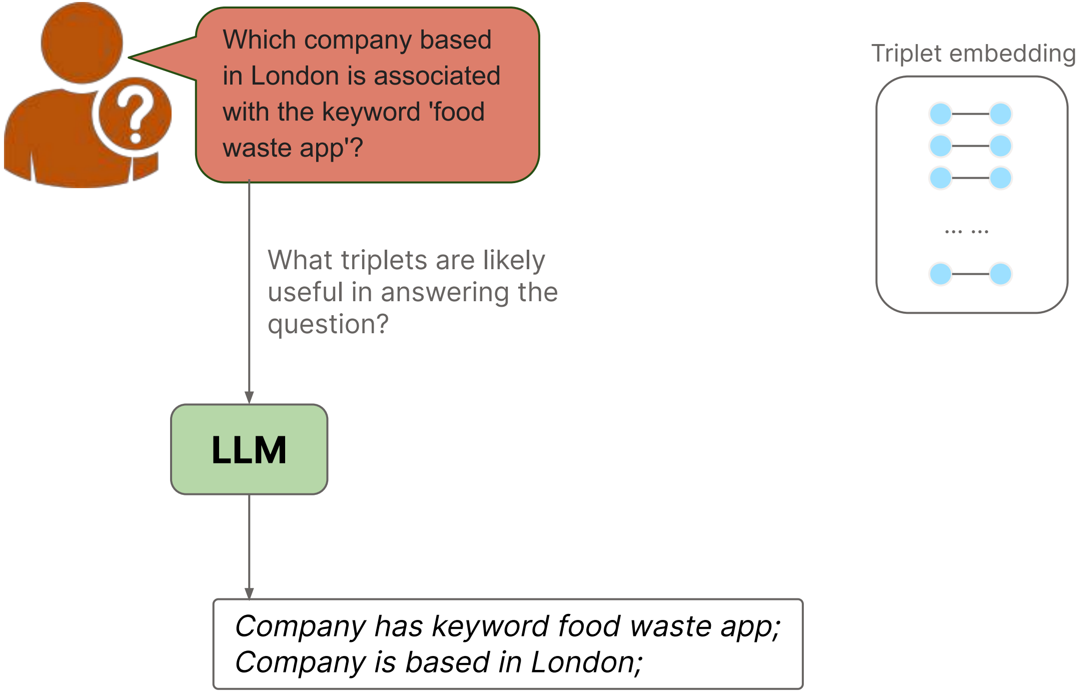
Contextual Retrieval: Semantic Triplets Search

Examples of “Triplet”s:

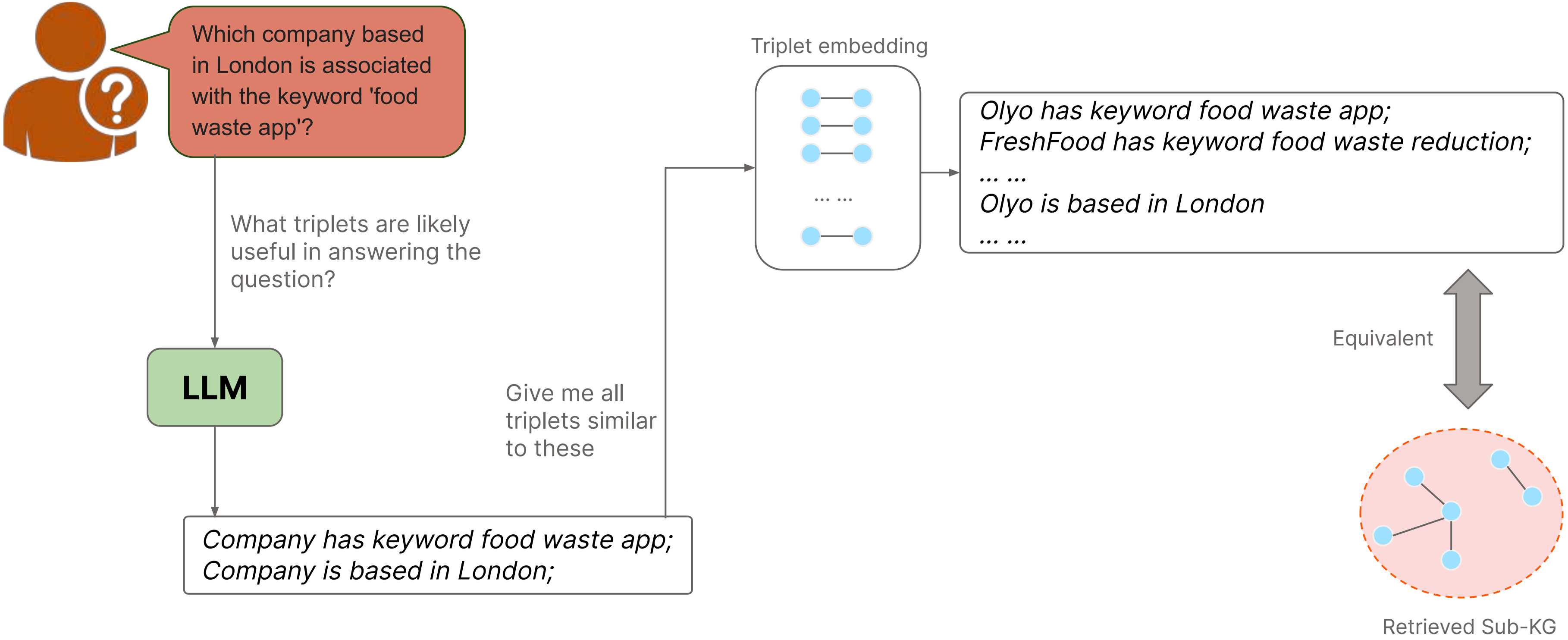
Start node	Relation	End node
EQT	— is based in	Stockholm
Martin	— worked with	Mark
…	…	…
KTH	— belongs to	University



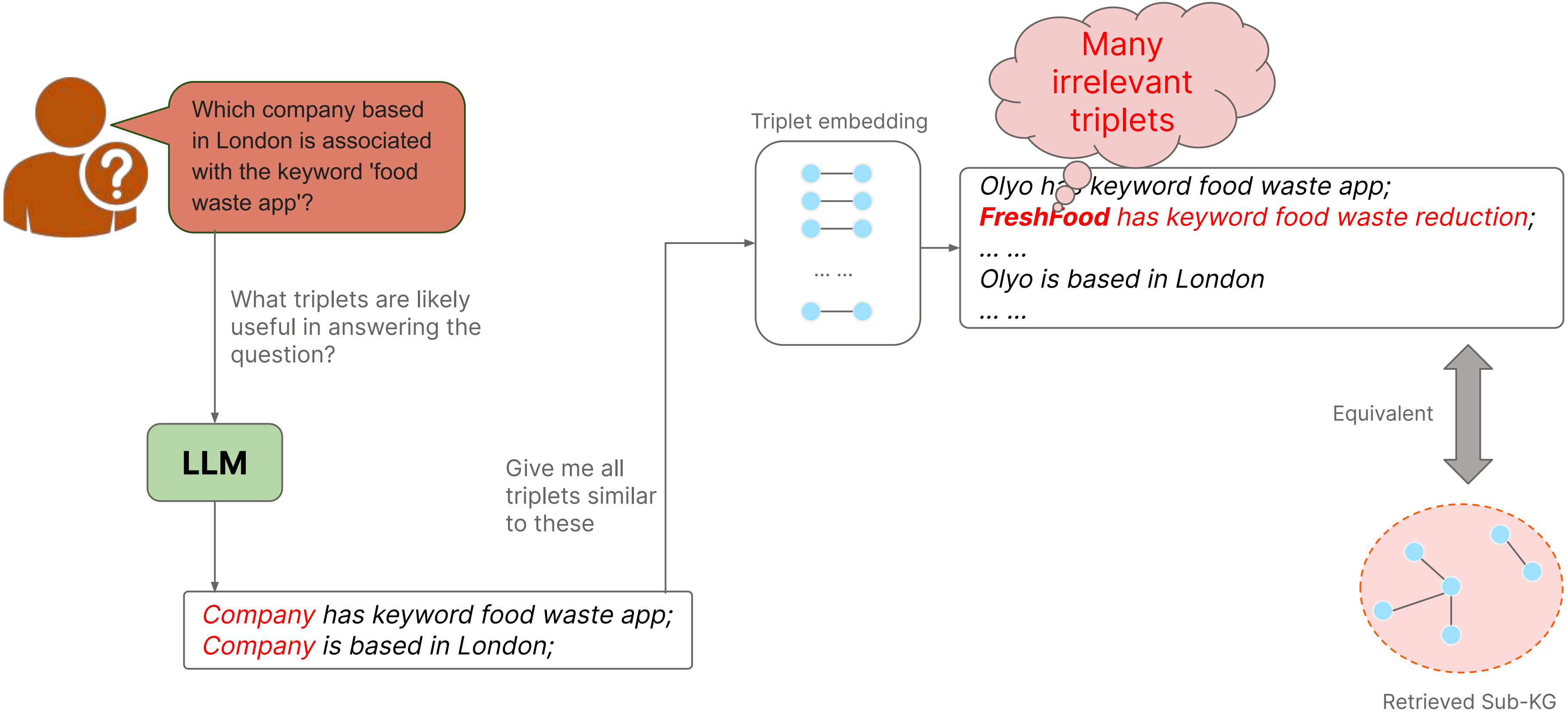
Contextual Retrieval: Semantic Triplets Search



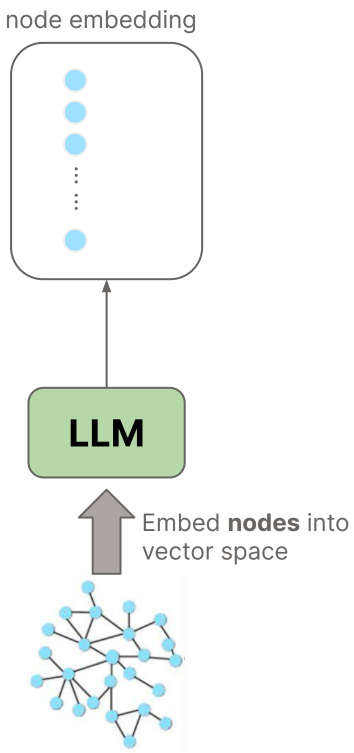
Contextual Retrieval: Semantic Triplets Search



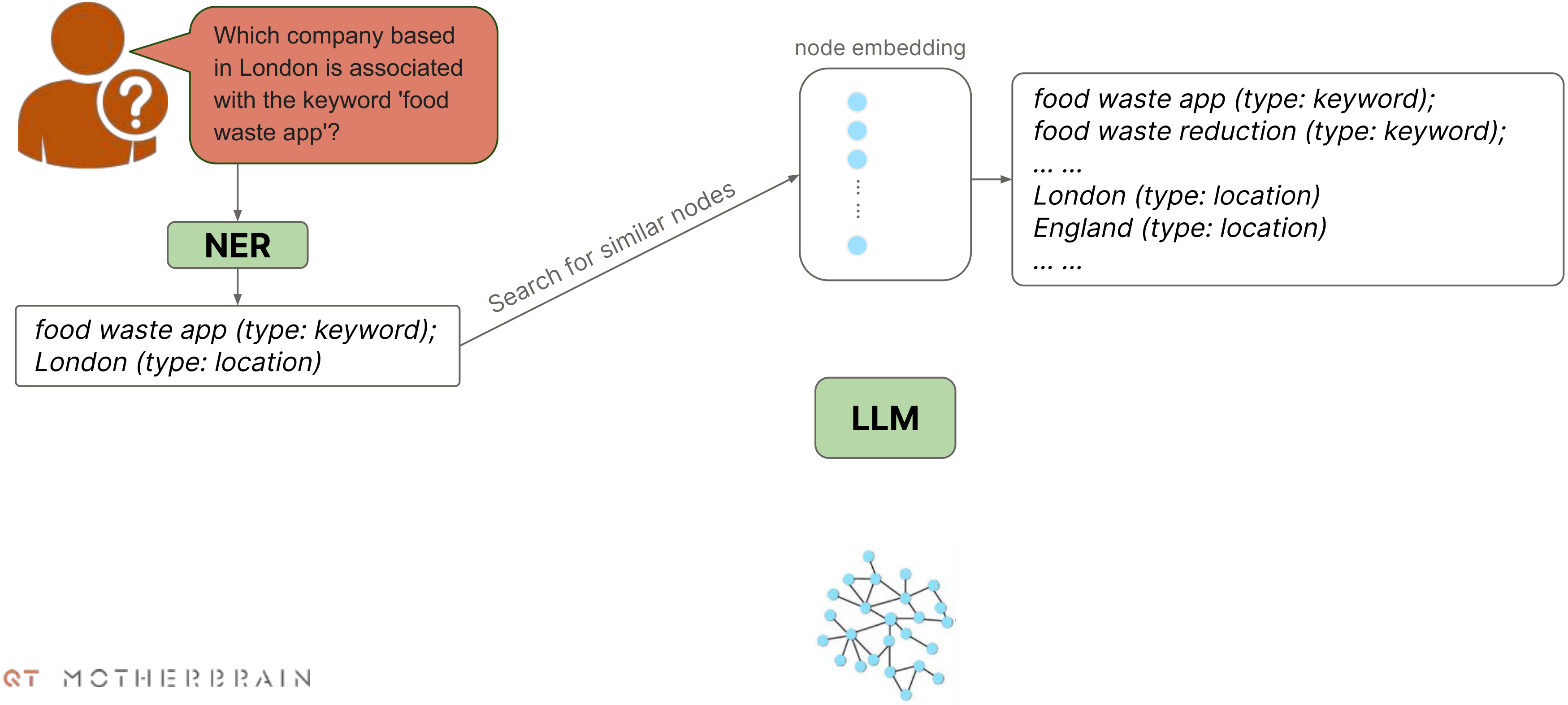
Contextual Retrieval: Semantic Triplets Search



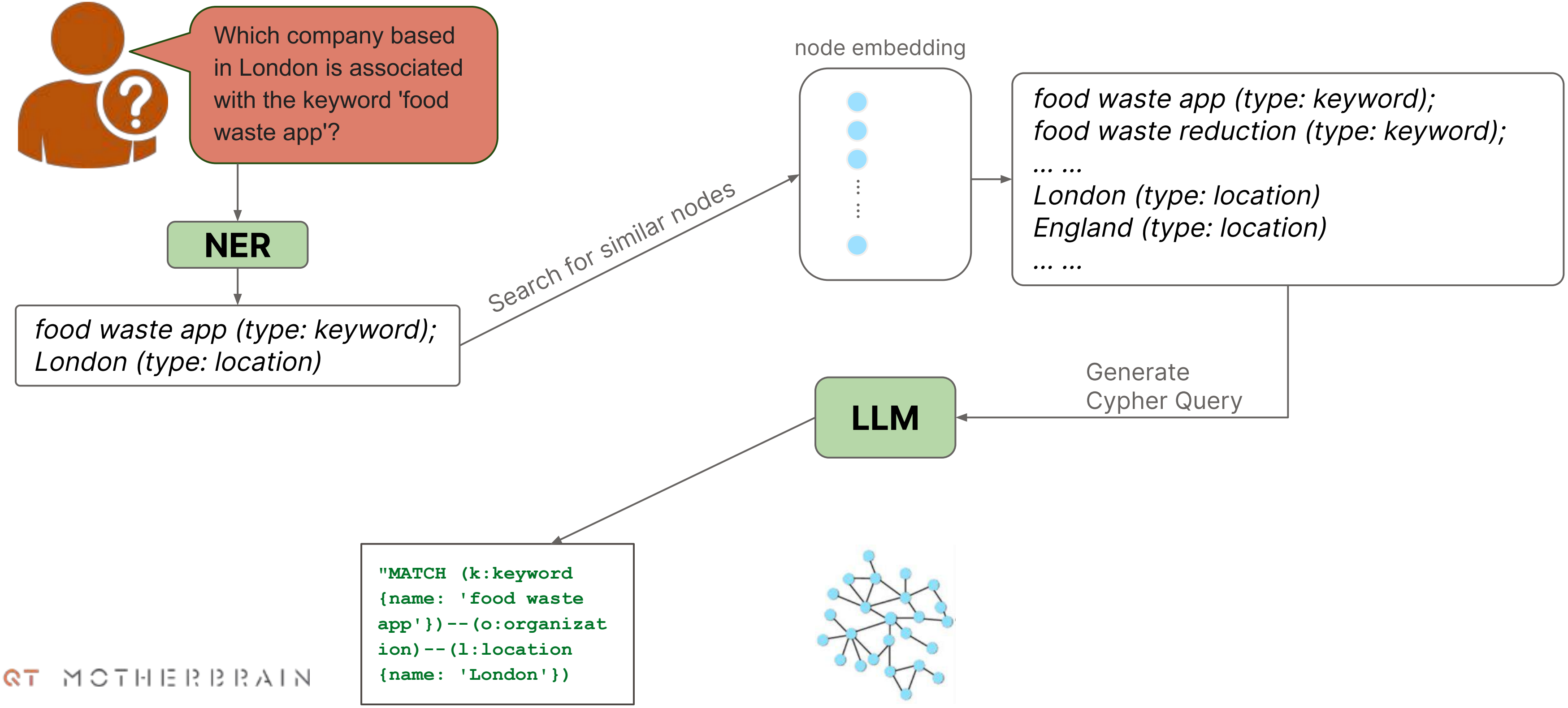
Contextual Retrieval: NER + Cypher



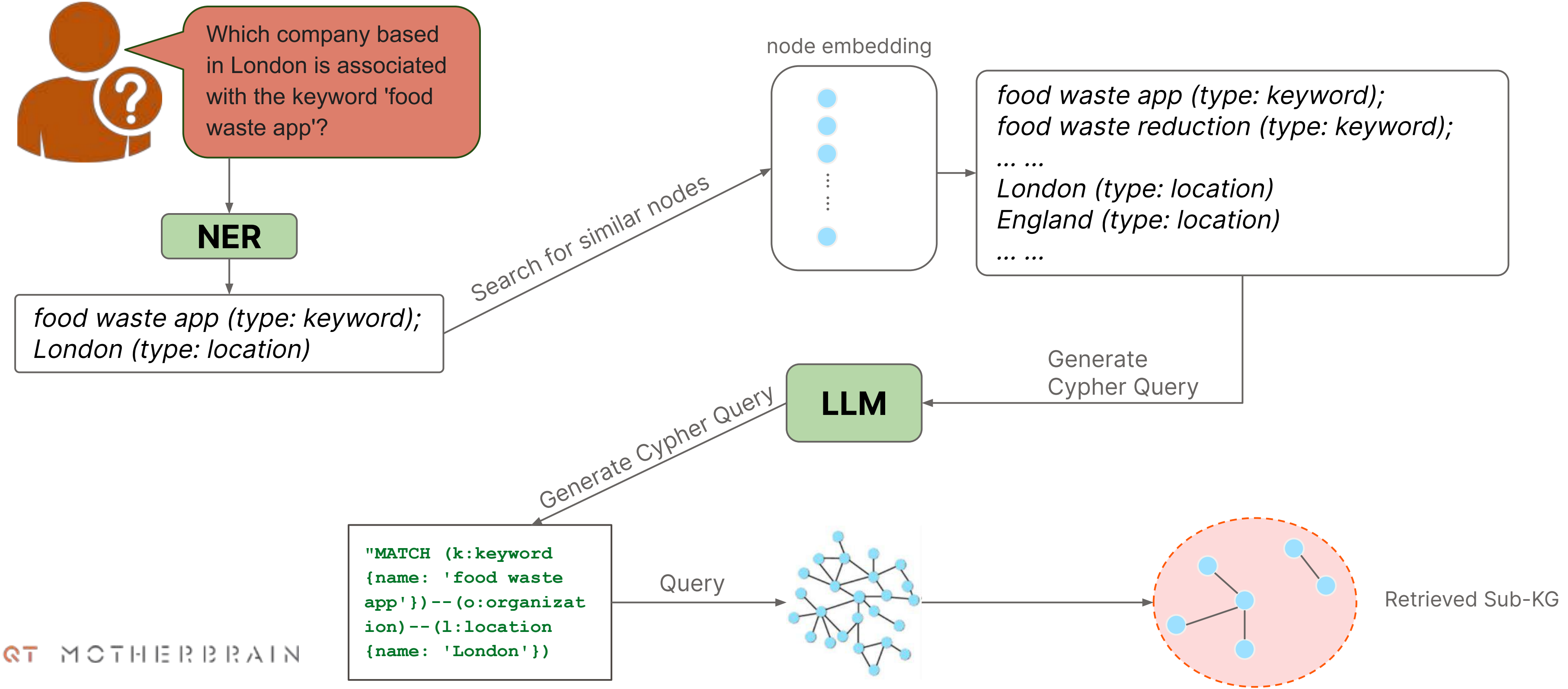
Contextual Retrieval: NER + Cypher



Contextual Retrieval: NER + Cypher



Contextual Retrieval: NER + Cypher



Agenda

Motivation

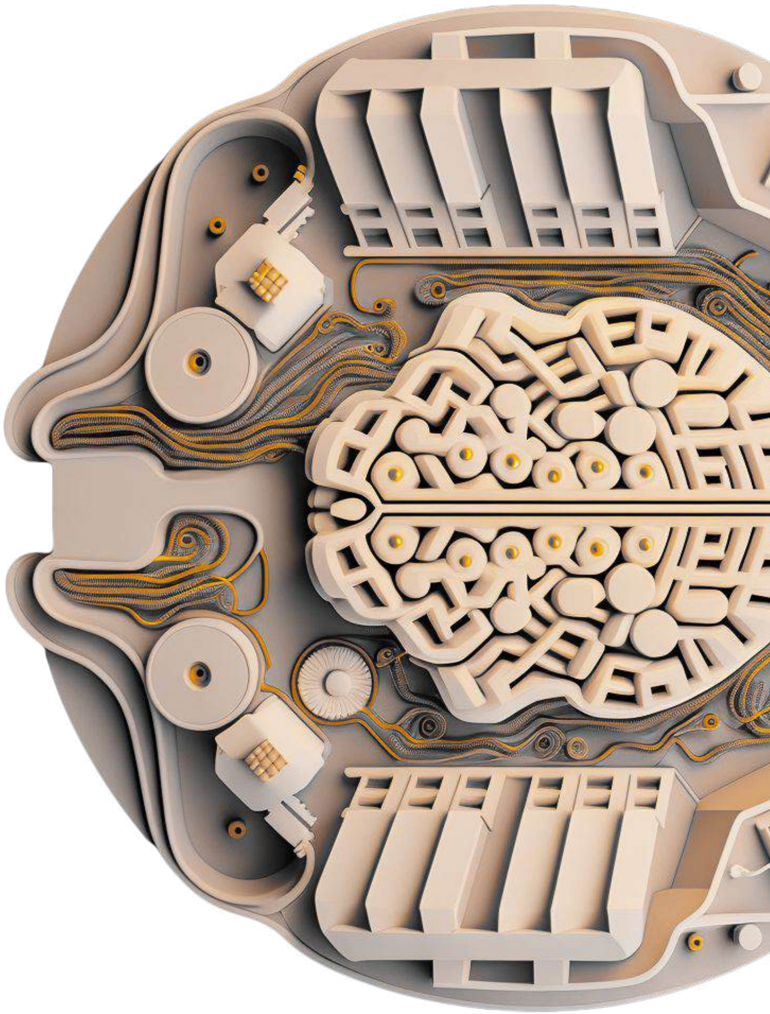
Approach

KG Construction

Contextual Retrieval

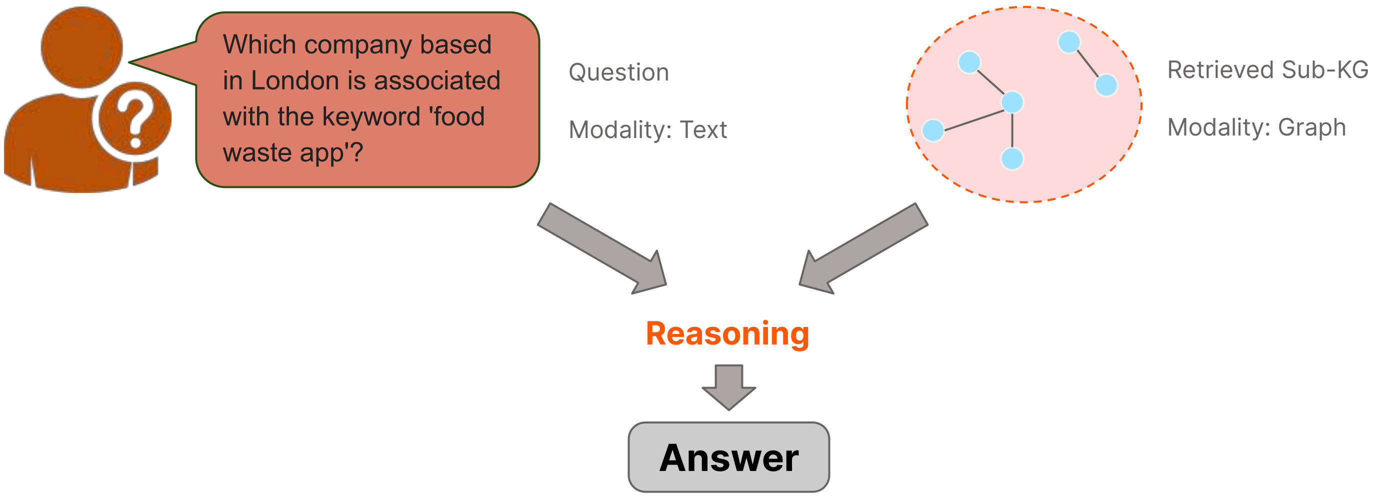
Reasoning

Summary



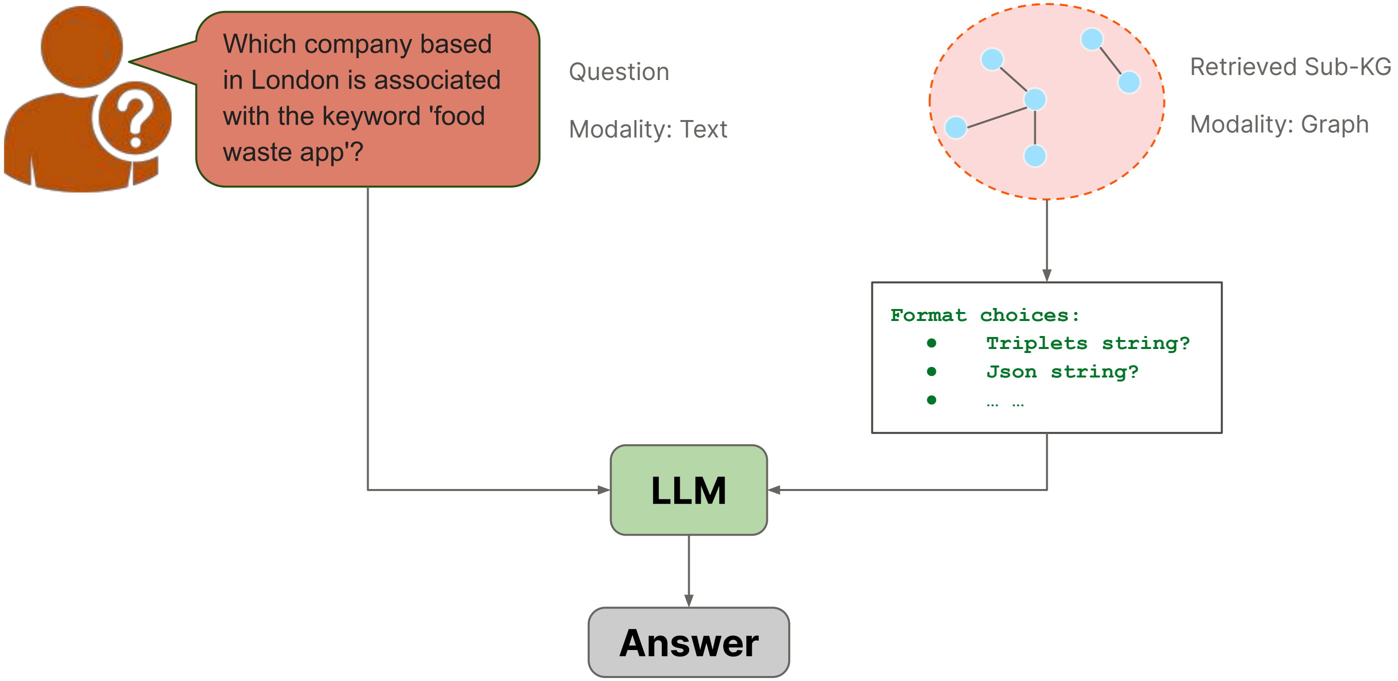
Reasoning

Objective: generate answer to the textual question using the retrieved sub-KG as input context.



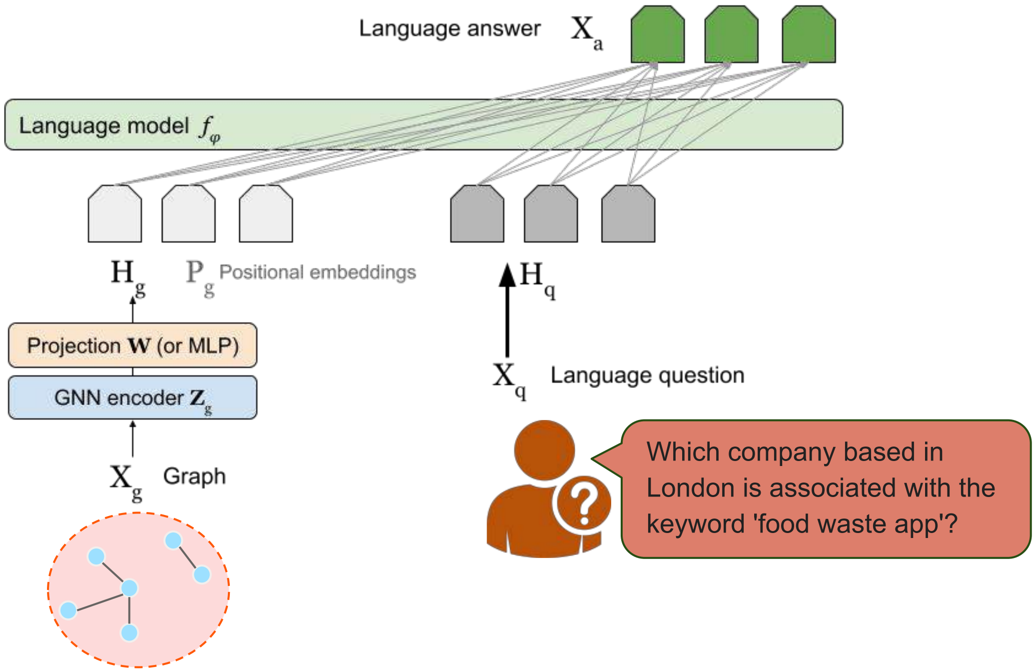
Reasoning: Zero-shot (simple option)

Zero-shot: transform the sub-KG into texts.



Reasoning: Cross-modal finetune (cool option)

Cross-modal finetune: jointly tune the word (text) and node (graph) into the same token space.



For more details, see this paper: https://arxiv.org/pdf/2402.07197

Agenda

Motivation

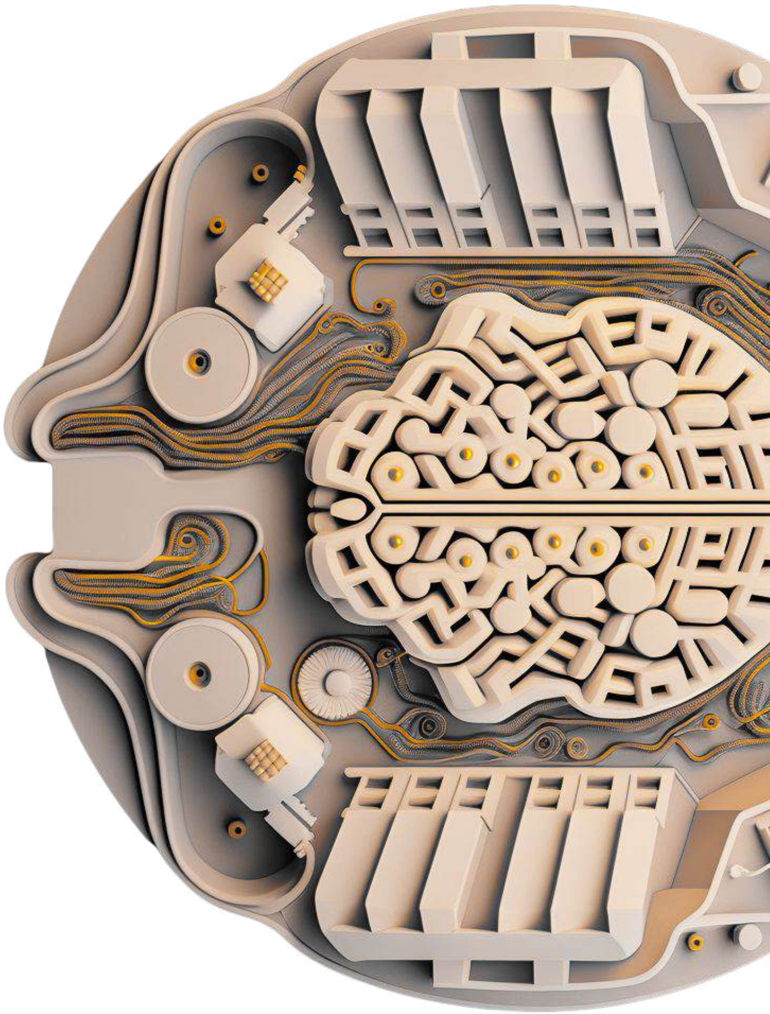
Approach

KG Construction

Contextual Retrieval

Reasoning

Summary



Summary

Knowledge Graph can empower LLM to facilitate PE investment activities.

Ask a question
We are looking at a deal of a company named ManyPets (manypets.com). Can you suggest some deal experts in related sectors and also some direct competitors in EU?

LLMs
ChatGPT
LLaMA
Gemini
Mistral
Claude
...

Reasoning

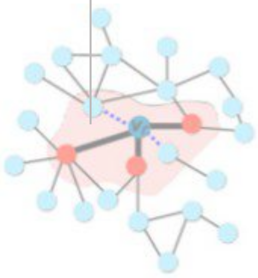
Answer
here are some suggestions for deal advisers in the pet insurance sector, as well as some direct competitors in the EU:
...

Retrieve relevant sub-graph

Third-Party Knowledge (Structured Data)

Construct

Proprietary Knowledge (Unstructured Docs)



Knowledge Graph

Summary

Knowledge Graph can empower LLM to facilitate PE investment activities.

Reasoning

Retrieve

- Define the entities and relation first.
- Combine proprietary and 3rd-party data.
- Multimodal LLM enables automatic KG extraction.

Construct

Summary

Knowledge Graph can empower LLM to facilitate PE investment activities.

Reasoning

Retrieve

- Identify named entities in question first.
- Graph database works better than vector database.
- Ask LLM to generate retrieval query.

Construct

- Define the entities and relation first.
- Combine proprietary and 3rd-party data.
- Multimodal LLM enables automatic KG extraction.

Summary

Knowledge Graph can empower LLM to facilitate PE investment activities.

Reasoning

- Zero-shot is simple and performs OK!
- Cross-modal finetune is new and cool, yet will be challenging to do well.

Construct

- Define the entities and relation first.
- Combine proprietary and 3rd-party data.
- Multimodal LLM enables automatic KG extraction.

Retrieve

- Identify named entities in question first.
- Graph database works better than vector database.
- Ask LLM to generate retrieval query.

Thanks!

- For further interaction, feel free to ping me on LinkedIn: www.linkedin.com/in/caolele
- Or, email us: tech_motherbrain-research@eqtpartners.com
- Learn more about EQT Motherbrain at:
  https://eqtgroup.com/motherbrain
  https://motherbrain.ai/