Graph RAG

## **Graph RAG**

# **Motivation**

为主流方案。而 GraphRAG 作为一种结合图结构与文本语义的检索增强方 法,具有结构化信息处理和多跳推理的潜力。本次调研主要围绕两个问题展 开: 1. GraphRAG 作为向量数据库能否输出结构化数据:传统向量数据库更偏 向于"黑箱式"的 Top-K 检索,而 GraphRAG 是否具备结构化输出能力

在当前大语言模型(LLM)驱动的问答系统中,检索增强生成(RAG)已成

(如实体、关系、社区) 将决定其在知识组织和追溯中的适用范围。 2. GraphRAG 的图关系处理流程是否适用于人物关系分析:人物关系往往 具有复杂的多跳连接、角色依赖和社群结构,GraphRAG 能否以结构化

Retriever

Retrieve | Return

方式刻画并支持推理,是评估其在情报研判、社交网络分析等场景中的 关键。 GraphRAG survey Discrimination Heuristic-based Learning-based Reranking Verbalization LLM-based Entity Linking Linear-based Pruning Name Entity Recognition nal Extr

Generator

Answer

**Explicit Construction** 

Implicit Construction

Deep Graph Encoder

(c) Shallow and Deep Graph Embedding

Global Answer

Graph Communities

Query Time

Reasoning Planning Graph 90 (P) Figure 3: A holistic framework of GraphRAG and representative techniques for its key components.

Query Decompositi

**Graph Data Source** 

GraphRAG 在 RAG 框架基础上引入图结构建模,整体机制包括两个核心阶 段: 1. 索引阶段 (Indexing) • 文本被分块 (chunking) 成较小的单元 (TextUnits) ; • 从文本中提取实体与关系,构建为图结构; 文本块与图结构分别进行嵌入,供后续查询使用; 图的节点/边与文本单元相链接,实现图-文联合索引。 2. 查询阶段 (Querying)

What is the nationality

查询被结构化或嵌入后用于匹配; 图结构被用于执行多跳检索,挖掘关联路径; 匹配到的图子结构 (子图) 与文本块一并提供给生成模型; 最终生成包含结构信息的回答。 Optimizing the binding affinity  $\leq -4.9 \text{ kcal/mol}$ Node Graph Edge of Jack 's wife ? 1 Entity Linking Molecule Graph Data Source

(b) Graph Kernal and Domain Expertise

city of 3 Graph Traversal (a) Entity Linking, Relational Matching, and Graph Traversal Figure 5: Visualizing representative retrievers used in GraphRAG.

Microsoft GraphRAG

Source Documents

Knowledge Graph

Indexing Time

取和模块化能力,主要流程包括:

索引阶段的关键模块:

查询阶段:

VS 区别与优势:

构);

Summary);

**Index Workflow** 

University of locate in

🔍 与传统RAG比较优势: 传统 RAG 通常基于文本块的相似度检索,忽略 实体之间的结构信息。GraphRAG的引入让检索具备"语义 + 图结构"双重约 束,尤其在多跳推理、实体对齐、因果链追踪等任务中具有天然优势。 <u>微博数据集,社会影响力局部性:建模转发行为 | AMiner --- Social</u>

<u>Influence Locality for Modeling Retweeting Behaviors | AMiner</u>

text extraction query-focused and chunking summarization Community Answers Text Chunks domain-tailored query-focused summarization summarization Entities & Relationships Community Summaries domain-tailored domain-tailored summarization , summarization community

detection

Pipeline Stage

三、Microsoft GraphRAG 的索引与查询机制及其改进点

微软开源的 GraphRAG 版本在传统 GraphRAG 基础上进一步增强了结构提

 文本分块(Chunking) → 文本单元(TextUnits) 2. 实体/关系抽取 → 图结构 (三元组形式) 3. 声明 (Claim) 提取 → 高阶语义信息 4. 社区发现 → 聚合相关节点构建"群体"

5. 图与文的多重嵌入 → 支持图、文本、社区的多模态检索

• 可利用图嵌入与语义嵌入联合进行召回;

• 支持子图级别的检索结果组织;

• 可基于图路径信息构建答案因果链。

Sample Workflow DAG

Prepare

Chunk

**EmbedDocuments** 

EmbedGraph

title

JD VANCE

DONALD

**TRUMP** 

type

**PERSC** 

**PERSC** 

**PERSC** 

**EVENT** 

**ORGAN** 

target

DONALD

**TRUMP** 

**MAGA** 

**MOVEMENT** 

HILLBILLY

**MIDDLETOV** 

YALE LAW

Phase 2: Graph Extra

**Entity & Relationship** 

Extraction

**Entity & Relationship** 

Summarization

Claim Extraction

**Graph Tables** 

SCHOOL

**ELEGY** 

OHIO

JD

VANCE

**Dataflow Overview** 

• Microsoft GraphRAG 明确输出结构化结果 (表格、社区、UMAP 结

支持多阶段数据流(如 Text → Graph → Community →

• 更适合用于复杂知识工程、情报挖掘和多跳问答任务。

### GenerateReports 提取实体 id human\_readable\_id 9cc787a9-0 0

2713-4b19-

b5b92777-

b01b-445d-

83b86ea5b865

a240-

bf05-

4

38579b8d13a0

0cfb2b8f-

f8f9-4a2b-

d828e8df525d

b5de-

**Dataflow Overview** 

**Documents** 

Phase 5: Document

Drockecika

Link to TextUnits

**Documents Table** 

Phase 6: Network

**Graph Embedding** 

**Umap Entities** 

**Final Entities** 

9636e5ad9238

8fe8-

1

ExtractGraph

**EmbedEntities** 

1



#### 0 118ff68d-0 JD VANCE e9c6-4dfb-853cf658e03ee46f 1 3ba9bd73-1 JD VANCE a908-4426b8f1c95263a983ca 2 700d6acb-2 JD VANCE 09eb-4cc8-845b-9ba8d7ee78fe 3 a482a176-3 JD VANCE d94f-4f4c-

4

Phase 1: Compose **TextUnits** 

Chunk

**Text Units** 

#### Phase 7: Text Embeddings Text Embedding **Description Embedding** GraphRAG 输出结构化数据:人物关系图谱能力探索 GraphRAG 不仅支持从文档中提取答案,还能结合结构化索引数据实现复杂 语义对象(如社交图谱、知识网络)的多模态表达。在本项目中,基于 GraphRAG 的 search\_engine.search() 方法进行局部语义检索,可获得 以下几类结构化数据: 1. 实体实体识别与聚合 (Entities) 返回结果中的 entities 字段提供了与查询相关的命名实体列表,字段包 括: id: 实体编号 • entity: 实体名称 (如"J.D. VANCE"、"DONALD TRUMP") • description: 实体说明 number of relationships: 图谱中该实体的连接数 in\_context: 当前查询是否包含该实体 这类信息为构建结构化知识图谱提供了节点基础。 2. 社交关系提取(Relationships) GraphRAG 会根据索引数据自动提取相关的双实体关系结构,字段包括: source / target: 关系的起始与目标实体 description: 自然语言描述关系内容 weight: 关系重要性 (可用于排序或上下文优先级评估) in\_context: 是否与当前查询显著相关 description id source target weight 112 U.S. SENATE 8.0 0 VANCE Vance serves as a senator in the U.S. Senate, ... **VANCE** 1 97 **HOWIE** Vance has 7.0 **BEIGELMAN** worked closely with Howie Beigelman ... HALF ISLAND Half Island 4.0 2 114 **VANCE** TV TV published commentary on Vance's... 3 35 **DONALD** J.D. Vance J.D. 27.0 **VANCE TRUMP** and Donald Trump are prominent figu... 50 J.D. VANCE 4 DONALD Donald 17.0 **TRUMP** Trump and J.D. Vance share a complex po... 41 J.D. **UKRAINE** J.D. Vance 7.0 68

### 构建和索引图数据库是 Grap GraphRAG 的性能。图数据: 谱。 实体别名

**Prompt Tuning** 

| index |  |
|-------|--|
|       |  |

--root ./ragtest \

--root ./ragtest \

--language Chinese \

--query "Vance的社交人物关系"

python -m graphrag prompt-tune \

--config ./ragtest/settings.yaml \

--domain "Social Relationship Analysis"

1/1

--method local \

| graphrag | init - |
|----------|--------|
| graphrag | index  |
| graphrag | query  |

**VANCE** 

J.D.

J.D.

J.D.

J.D.

VANCE

**VANCE** 

**VANCE** 

**VANCE** 

**TAIWAN** 

**MIDDLETOWN** 

**KENTUCKY** 

**OUR OHIO** 

**RENEWAL** 

69

70

71

72

42

60

74

75

| 的多条明确互动关系,涉及政治联盟                              |
|---|
| 建和索引图数据库是 GraphRAG 的基<br>iphRAG 的性能。图数据来源主要包护 |
| <b>体别名</b>                                    |
| ompt Tuning                                   |
| graphrag initroot ./ragte                     |
|   |
| graphrag indexroot ./ragt                     |
|   |
| graphrag query \root ./ragtest \              |

例如,对于"Vance的社交关系"查询,系统输出了其与 Donald st est

J.D. Vance's grandparents migrated from Kentuc... J.D. Vance founded Our Ohio Renewal to address...

has strongly

J.D. Vance's

stance on

aligns with

J.D. Vance

and raised in

Middletown,

was born

Taiwan

Dona...

6.0

25.0

7.0

6.0

opposed providing

war ...

Trump、Howie Beigelman、Yoram Hazony、Mitt Romney 等人物 政治联盟、宗教合作、思想传承等维度。 [础,其中图数据库的质量直接影响 括两类: 开放知识图谱和自建知识图 SHELL SHELL **SHELL** 

SHELL