Large Language Models Meet Knowledge Graphs for Question Answering: Synthesis and Opportunities

Chuangtao Ma, Yongrui Chen, Tianxing Wu, Arijit Khan, Haofen Wang

Aalborg University, Southeast University, Bowling Green State University, Tongji University

Motivation

- Complex question answering (QA) usually involves knowledge interactions and fusion across modalities and sources.

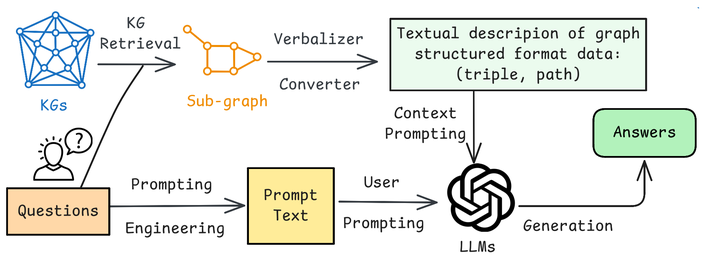

Fig. 1: KG-RAG for QA.

- LLM and naive retrieval augmented generation (RAG) suffer significantly from the lack of factual knowledge and limited reasoning capability.

KGs as Reasoning Guidelines

- KGs can provide reasoning guidelines for LLMs to access precise factual evidence chains.
- **Challenge**: *How to improve the reasoning efficiency and capabilities over the large-scale graph and incomplete KGs?*

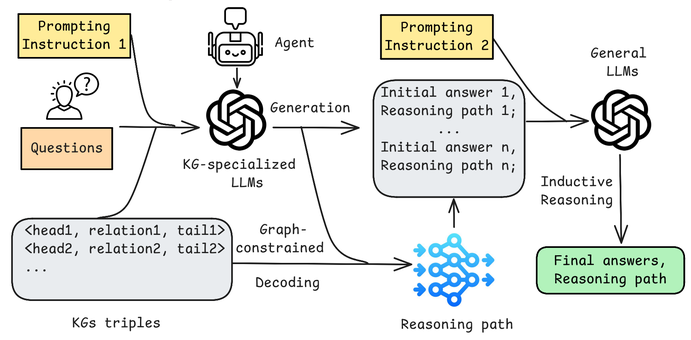

Fig. 5: KGs as Reasoning Guidelines.

KGs as Refiners and Validators

- KGs act as refiner and validator for LLMs, where the factual evidence from KGs enables LLMs to refine and verify the intermediate answers.
- **Challenge**: *How to ensure knowledge in KGs is up-to-date and correct, and handle the knowledge conflict between intermediate answers and KG factual knowledge?*

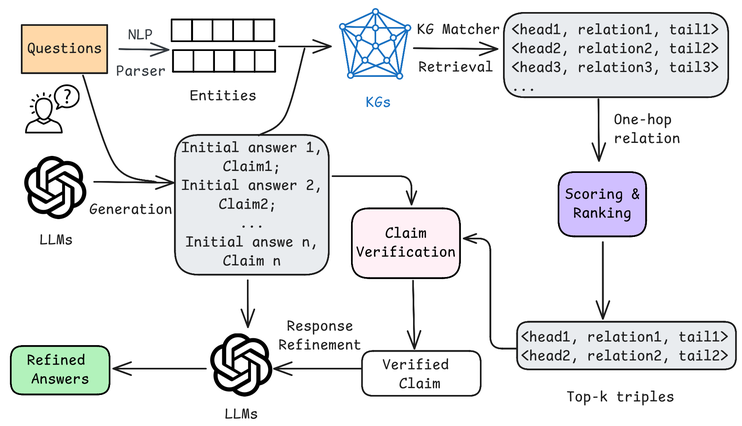

Fig. 6: KGs as Refiners and Validators.

Taxonomy and Alignment

- **Taxonomy**: A structured taxonomy from different perspectives aims to highlight the alignments between various LLM+KG approaches and complex QA by discussing how LLM+KG approaches with different roles for KG can address the challenges of complex QA.

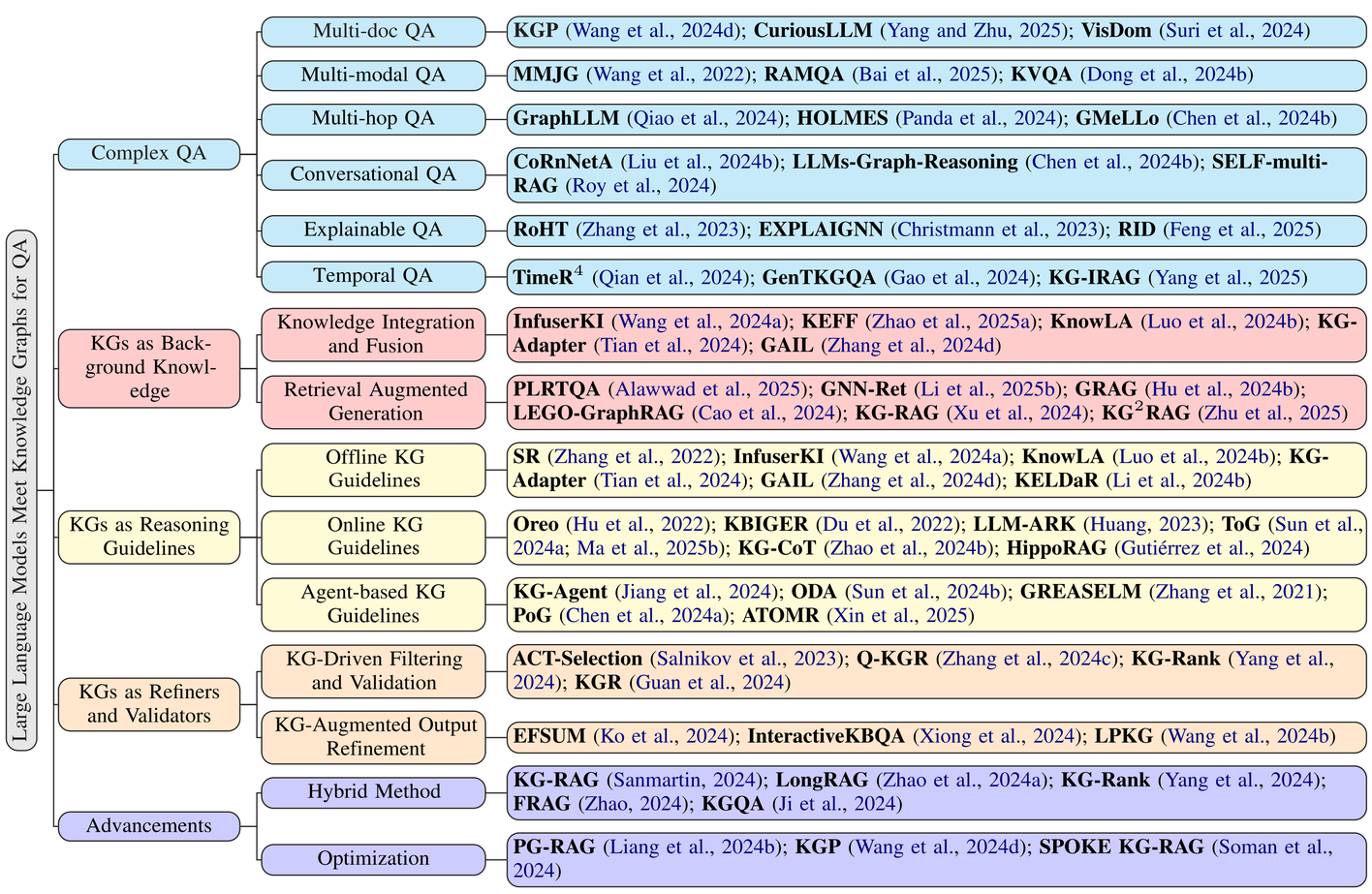

Fig. 2: A Structured Taxonomy of Synthesizing LLMs and KGs for QA.

- **Alignment**: The alignment of existing approaches in synthesizing LLMs and KGs for complex QA.

Table 1: Research Progress on synthesizing LLMs and KGs for complex QA. ✔ Fully Investigated. △ Partially Investigated. ✕ Not Yet Investigated.

Approach	Multi-doc QA	Multi-modal QA	Multi-hop QA	Conversational QA	XQA	Temporal QA
KG as Background Knowledge	✔	✔	✔	✔	△	✔
KG as Reasoning Guidelines	✔	✔	✔	✕	✔	✔
KG as Refiners and Validator	△	✕	△	✔	✕	△
Hybrid Methods	✔	✔	✔	✔	✔	✔

Advacements

- The hybrid method can address the challenges of complex QA in knowledge interactions and reasoning across modalities and knowledge sources.

Table 2: Comparison of Hybrid Approach.

Methods	Techniques	KG(s)	QA Type	Metric(s)
LongRAG [2024]	Domain-specific Fine-Tuning for RAG & CoT-guided Filter	Wikidata	KBQA, Multi-hop QA	F1
SimGRAG 2025]	Instruction Fine-Tuning for RAG with Filter	Wikipedia, PubMed	Domain-specific QA, Multi-choice QA	Acc, Rouge-L, MAUVE, EM, F1
KGQA [2024]]	KG-related Instruction Tuning with CoT Reasoning	Dataset Inherent KGs	KGQA	Hits@1, F1
KG-IRAG [2025]]	Incremental Retrieval and Iterative Reasoning	Self-constructed KGs	Temporal QA	EM, F1, HR, HAL
PIP-KAG [2025]]	Parameteric Pruning for KAG	Dataset Inherent KGs	KGQA, Multi-hop QA	EM, ConR, MR
RAG-KG-IL [2025]	Agent-based Incremental Learning and Knowledge Dynamic Update	Self-constructed KGs	Domain-specific QA	AM, HVC
CoT-RAG [2025]	KG-driven CoT Generation and Knowledge-aware RAG with Pseudo-program KGs	Self-curated Pseudo-Program KGs	KGQA, Multi-hop QA	RA, Robustness

- Even if several optimizations have been investigated to reduce the costs of vector retrieval and reasoning, the relevant subgraphs retrieval and graph reasoning remain a costly task.
- **Challenge**:*How to achieve efficient vector indexing and search at scale KGs and maintain a balance between cost and performance?*

Takeaways

- **Efficiency-aware retrieval**. Current retrieval and prompting pipelines repeatedly query the KG for every search, while batch retrieval and caching subgraphs can reduce the retrieval cost.
- **Quantify knowledge alignment**. The absence of metrics scores the semantic overlap and structural compatibility between parametric knowledge of LLM and symbolic knowledge of KG.
- **Adaptive knowledge reasoning at scale**. Reasoning over a large graph and language model remains a costly task, while exploiting adaptive reasoning and incremental computation-friendly hardware can mitigate the reasoning cost.
- **Fairness-aware and explainable QA**. Developing conversational QA with retrieval strategies can dynamically detect and adjust knowledge biases and further improve the explainability of QA via multi-turn user interactions.

KGs as Background Knowledge

- **Knowledge Integration and Fusion** enhances LMs by fusing and training text and KG jointly, while **RAG** retrieves relevant knowledge and then augments LLMs with retrieved knowledge via a prompt.
- **Challenge**: *How to retrieve the relevant knowledge from large-scale KGs and then effectively integrate and fuse with LLMs without inducing knowledge conflict?*

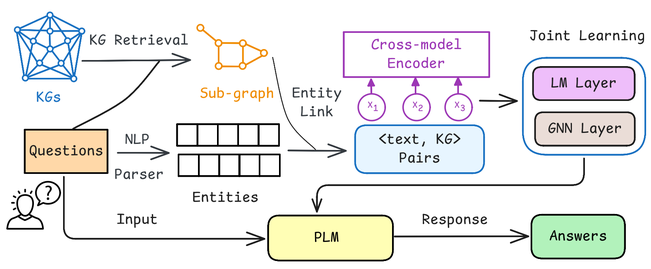

Fig. 3: Knowledge Integration and Fusion.

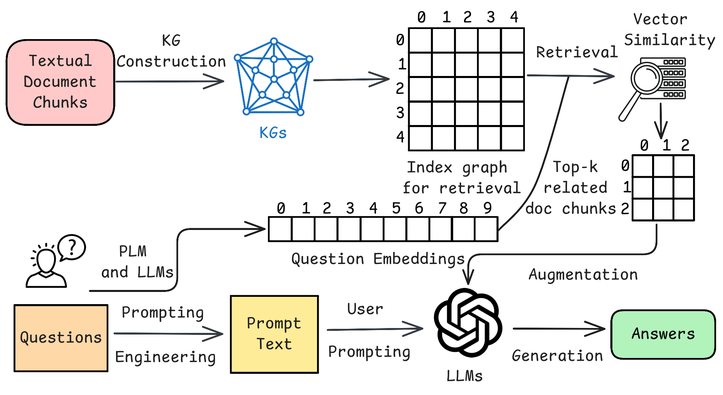

Fig. 4: Retrieval Augmented Generation.

Contact

- Chuangtao Ma, E-mail:chuma@cs.aau.dk