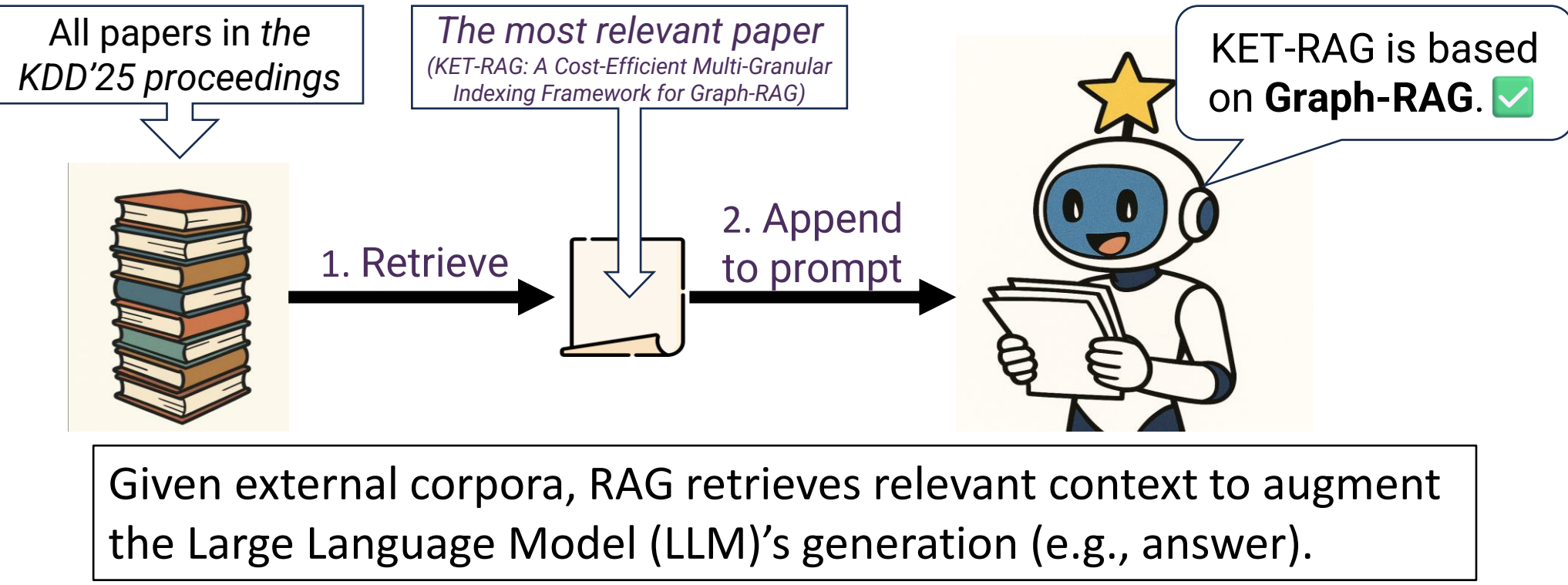
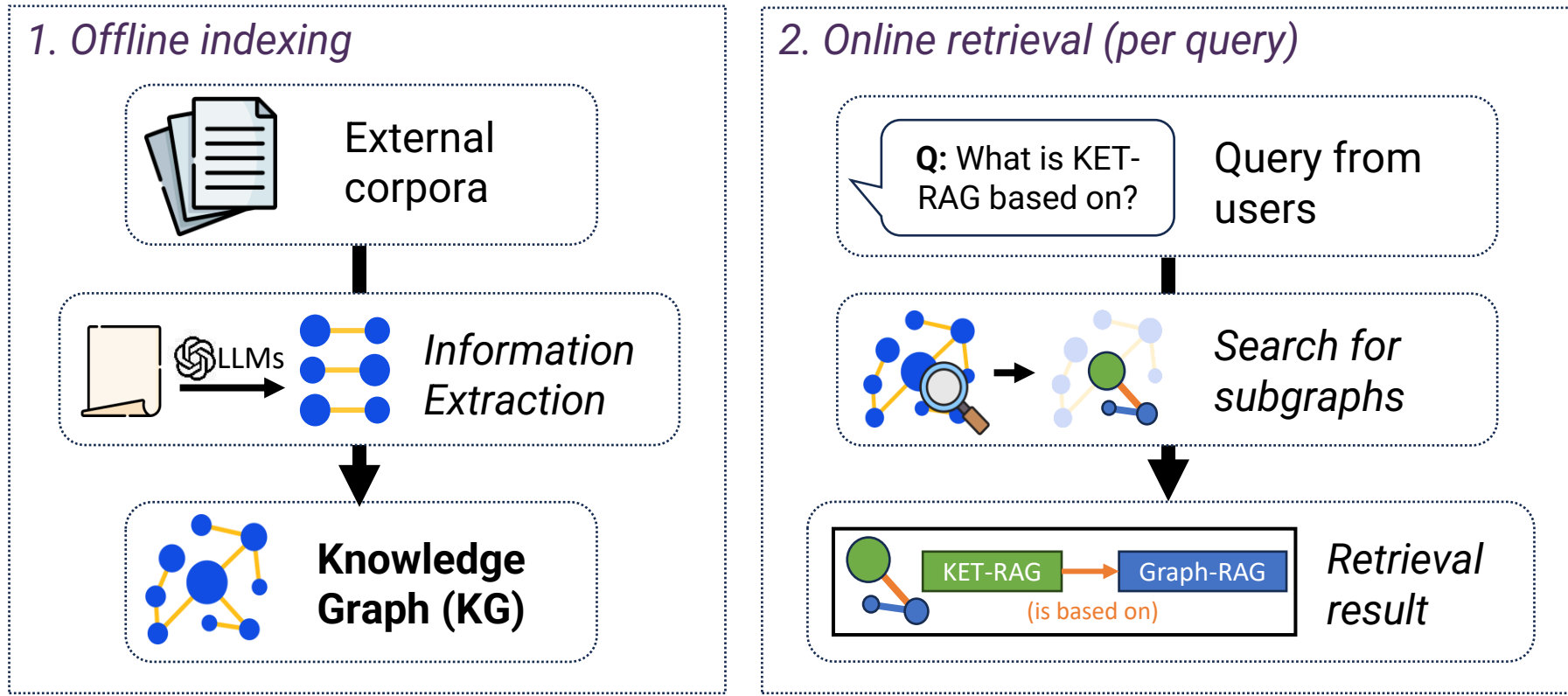


RAG and Graph-RAG

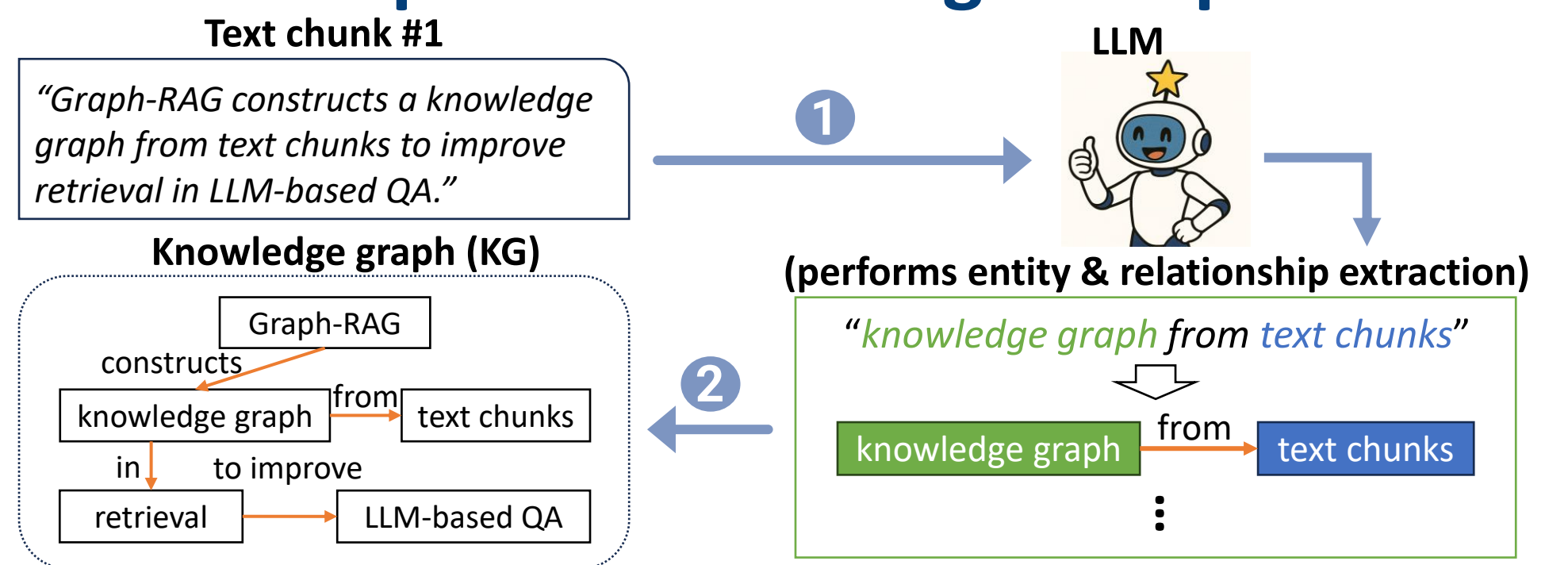
• RAG (Retrieval-Augmented Generation)



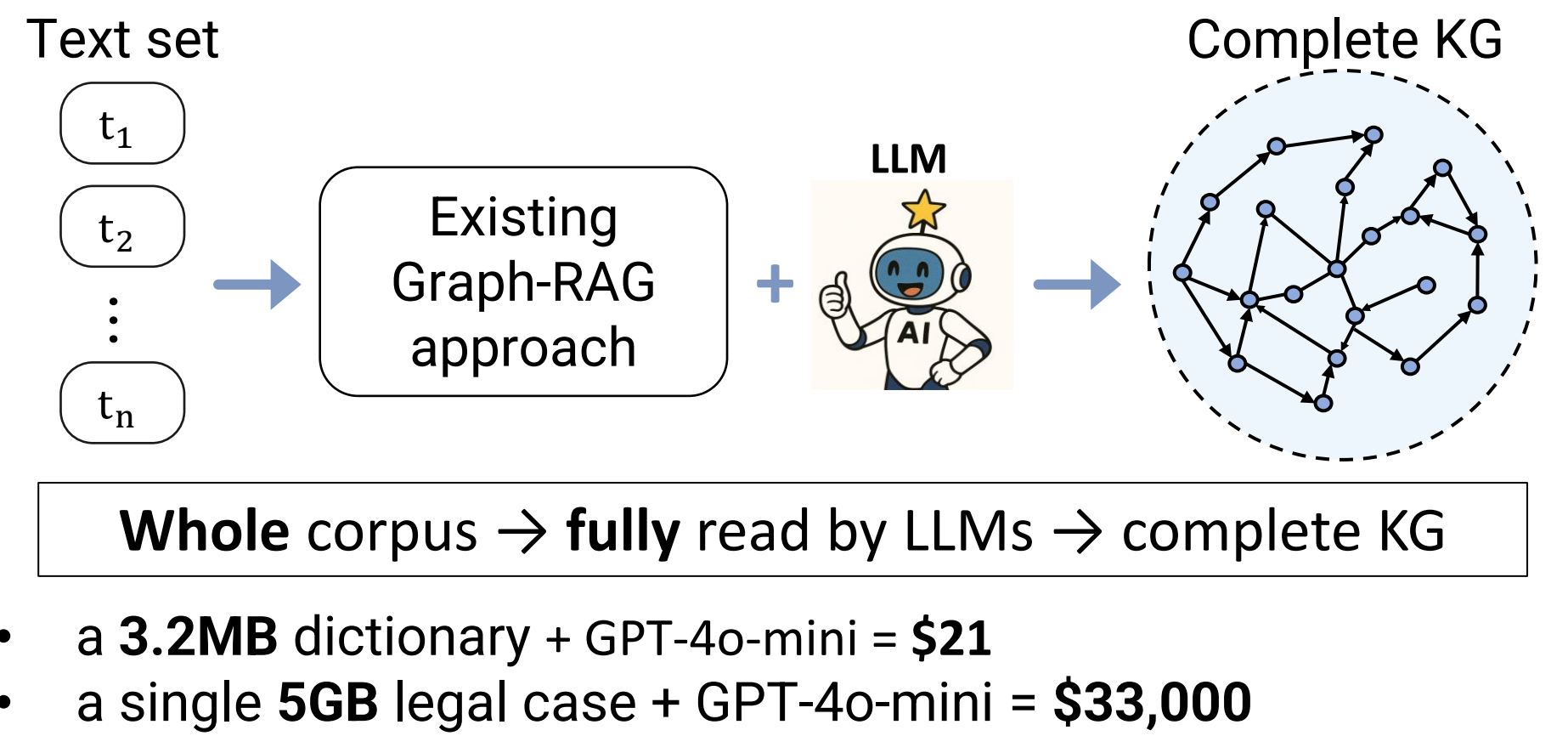
• Graph-RAG (Graph-based Retrieval-Augmented Generation)



Graph-RAG Indexing Example

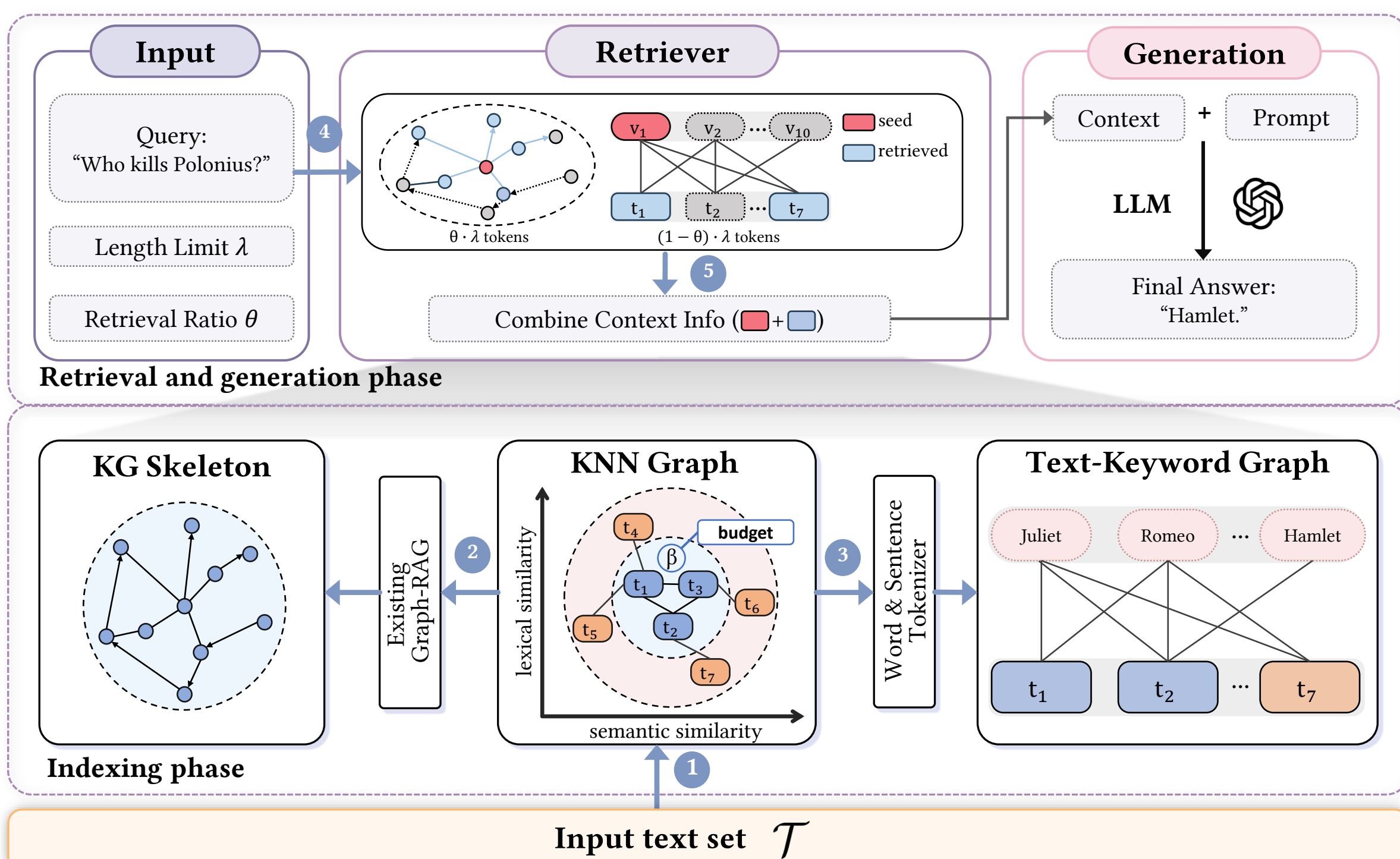


Limitations

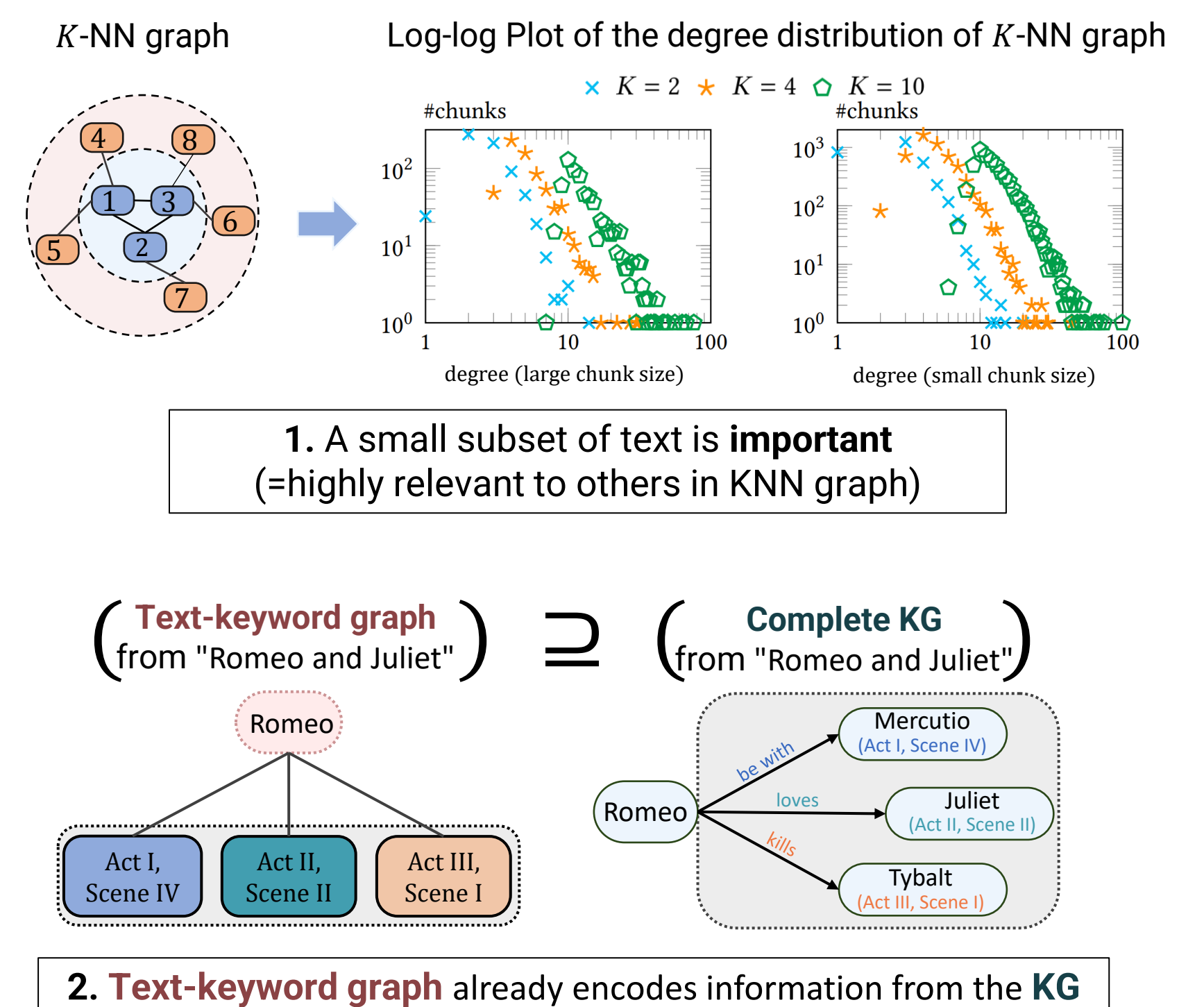


Our Solution: KET-RAG

• Framework overview



• Key design and rationale



Experiments

• Scenario

- Multi-hop QA (datasets: MuSiQue, HotpotQA)
- open-ended QA (datasets: RAG-QA Arena)

• Existing competitors

- (1) Vanilla RAG(Text-RAG) (2) KNN-Graph-RAG (3) Microsoft's Graph-RAG (4) Hybrid-RAG (5) HyDE (6) HippoRAG (7) LightRAG

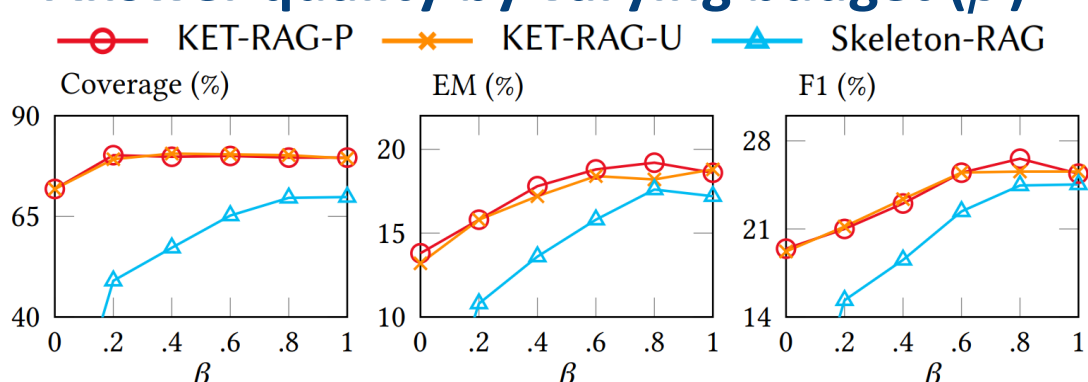
• Proposed baselines

- Keyword-RAG, Skeleton-RAG

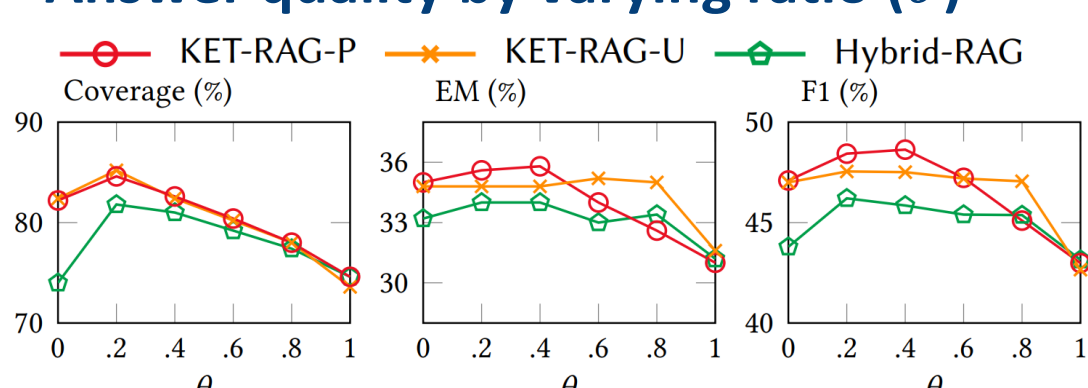
• Final solutions

- KET-RAG-U, KET-RAG-P

• Answer quality by varying budget (β)



• Answer quality by varying ratio (θ)



• Overall performance (best = bold, second-best = underlined)

- Main result: KET-RAG-U, KET-RAG-P** have superior performance and $\approx 80\%$ of the cost compared to Graph-RAG

Dataset	MuSiQue					HotpotQA					RAG-QA Arena	
	USD	Coverage	EM	F1	BERTScore	USD	Coverage	EM	F1	BERTScore	USD	Win Rate
Text-RAG	0.02	26.4	3.0	5.4	62.9	0.01	37.2	14.8	19.4	69.9	0.03	0.0
KNNG-RAG	0.02	21.2	2.8	4.6	62.5	0.01	28.6	13.4	16.9	68.8	0.03	17.8
MS-Graph-RAG	2.30	47.6	11.4	15.8	67.4	2.30	63.0	21.6	30.2	74.2	5.05	33.4
Hybrid-RAG	2.32	49.2	10.4	15.1	<u>68.8</u>	2.31	64.8	22.6	30.5	76.8	5.08	35.0
HyDE	0.03	33.0	4.2	6.7	63.5	0.02	40.8	16.2	21.3	70.9	0.03	33.2
HippoRAG	1.49	51.6	9.2	14.2	66.8	1.40	64.0	29.2	<u>38.3</u>	77.3	1.37	<u>38.4</u>
LightRAG-Local	1.80	39.0	9.4	12.9	66.6	1.77	57.6	22.4	28.1	73.1	4.00	29.0
LightRAG-Global	1.80	37.6	6.4	9.5	64.8	1.77	49.0	19.6	24.5	71.8	4.00	23.0
LightRAG-Hybrid	1.80	45.6	10.2	14.4	67.2	1.77	61.6	24.6	30.7	74.2	4.00	30.0
Keyword-RAG	0.03	50.8	7.0	11.8	68.1	0.03	60.2	24.4	33.5	78.9	0.07	37.0
Skeleton-RAG	1.86	43.4	11.0	14.1	66.8	1.84	57.8	20.0	26.7	73.2	4.04	28.6
KET-RAG-U	1.89	<u>76.2</u>	<u>13.4</u>	<u>18.1</u>	68.1	1.87	<u>81.4</u>	28.4	38.2	<u>77.5</u>	4.08	35.0
KET-RAG-P	1.89	77.0	14.0	18.9	69.0	1.87	81.6	<u>28.6</u>	38.7	<u>77.5</u>	4.08	39.6

Acknowledgement, technical report, code repository, etc.

This research is supported by the Ministry of Education, Singapore, under its MOE AcRF TIER 3 Grant (MOE-MOET32022-0001).

For the technical report and code, please scan the QR code and visit Yiqian's homepage.

