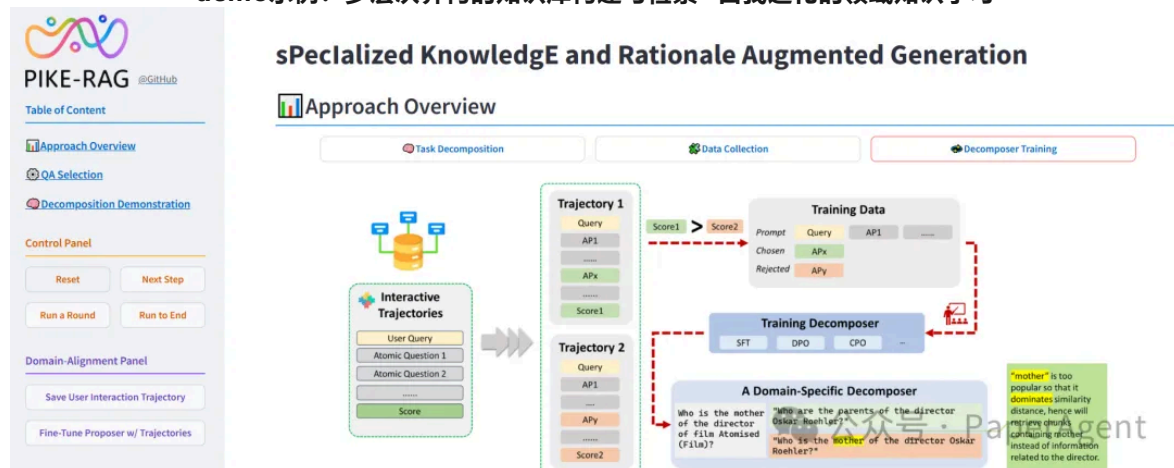


比GraphRAG更懂“思考”，微软又开源PIKE-RAG：主打复杂私域知识理解和推理

PaperAgent 2025年02月14日 11:26 湖北

继GraphRAG之后，微软又发布**PIKE-RAG**，主打在**复杂企业场景中私域知识提取、推理和应用能力**，PIKE-RAG 已在工业制造、采矿、制药等领域进行了测试，显著提升了问答准确率。**报告、代码、demo均已开源（在文末）。**

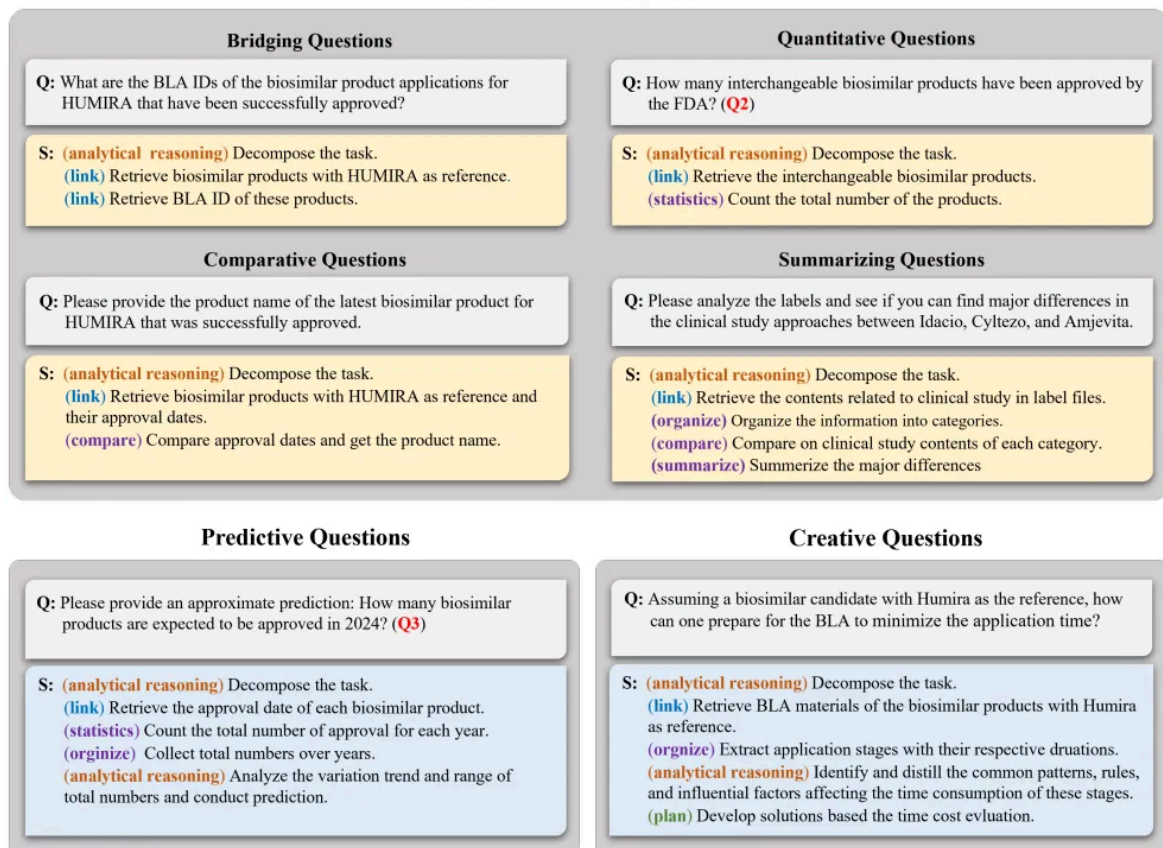
demo示例：多层次异构的知识库构建与检索+自我进化的领域知识学习



RAG系统在满足现实世界应用的复杂和多样化需求方面仍然面临挑战。仅依靠直接检索不足以从专业语料库中提取深度领域特定知识并进行逻辑推理。

企业场景复杂多样的Query

Linkable & Reasoning Questions

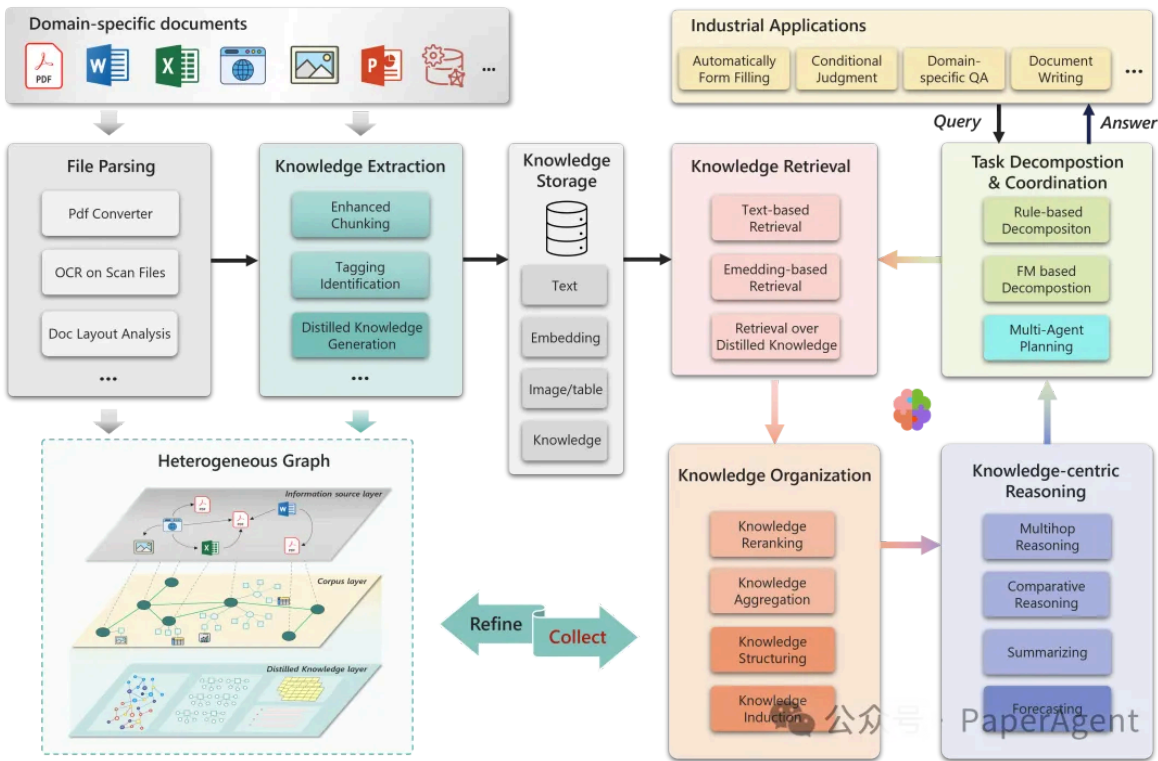


Answers explicitly present in corpus

Answers implicitly exist in corpus

Answers not present in corpus

基于此，微软亚洲研究院提出了 **PIKE-RAG** (sPecalized KnowledgeE and Rationale Augmented Generation) 方法，该方法专注于提取、理解和应用领域特定知识，同时构建连贯的推理逻辑，以逐步引导 LLM 获得准确的响应。



PIKE-RAG框架主要由几个基本模块组成，包括文档解析、知识抽取、知识存储、知识检索、知识组织、以知识为中心的推理以及任务分解与协调。通过调整主模块内的子模块，可以实现侧重不同能力的RAG系统，以满足现实场景的多样化需求。

Decomposition Demonstration

Round 4 [Answer Achieved]

Final Answer

Morning Edition ✓ CORRECT!

Rationale

The context provided states that Bob Edwards was the first host of National Public Radio's flagship program, "Morning Edition". It also mentions that Renée Montagne succeeded Bob Edwards as the co-host of "Morning Edition" along with Steve Inskeep and later David Greene.

Context Atomic Questions

	Context
0	Renée Montagne (pronounced Mon-TAIN) is an American radio journalist and was the co-host (with Steve Inskeep and David Greene) of National Public Radio's weekday morning news program, "Morning Edition", from May 2004 to November 11, 2016. Montagne and Inskeep succeeded longtime host Bob Edwards, initially as interim replacements, and Greene joined the team in 2012. Montagne had served as a correspondent and occasional host since 1989. She usually broadcasts from NPR West in Culver City, California, a Los Angeles suburb.
1	Robert Alan "Bob" Edwards (born May 16, 1947) is an American broadcast journalist, a Peabody Award-winning member of the National Radio Hall of Fame. He gained reputation as the first host of National Public Radio's flagship program, "Morning Edition". Starting in 2004, Edwards then was the host of "The Bob Edwards Show" on Sirius XM Radio and "Bob Edwards Weekend" distributed by Public Radio International to more than 150 public radio stations. Those programs ended in September 2015.
2	The Bob Edwards Show is an American radio program previously presented by Sirius XM Satellite Radio every weekday morning at 8 a.m. Eastern, with repeats at 8 a.m. Central, 7 a.m. Pacific, 6 p.m. Mountain, and the next day at 7 a.m. Eastern. The program was heard on the Sirius XM Public Radio station at XM channel 121 and Sirius channel 205, and was also available 24/7 on XM Radio Online and Sirius Internet Radio.

NOTE: The supporting facts are displayed in violet.

Current RoundRound 3Round 2Round 1

Atomic Proposal

ProposalsRationale

	Proposal
0	What radio program did Bob Edwards first host?
1	Who succeeded Bob Edwards as the host of that program?
2	When did Renée Montagne start co-hosting 'Morning Edition'?
3	Did Renée Montagne co-host any other program that Bob Edwards hosted?
4	What is the relationship between 'The Bob Edwards Show' and 'Morning Edition'?
5	Was 'The Bob Edwards Show' ever hosted by Renée Montagne?
6	What are the key differences between 'Morning Edition' and 'The Bob Edwards Show'?
7	What program was Renée Montagne associated with when Bob Edwards left 'Morning Edition'?

Retrieval

Select Result Index:
0

- Proposal Question (Index: 9)
What is the significance of NPR West in Culver City, California, in relation to Renée Montagne's broadcasting career?

- Relevant Atomic Question
Which television show did Craig Kilborn originally host?

- Relevant Chunk
Craig Kilborn (born August 24, 1962) is an American comedian, sports and political commentator, and television host. He was the original host of "The Daily Show", a former anchor on ESPN's "SportsCenter", and Tom Snyder's successor on CBS' "The Late Late Show". On June 28, 2010, he launched "The Kilborn File" after a six-year absence from television. "The Kilborn File" aired on some Fox stations during a six-week trial run. In comedy, he is known for his deadpan delivery.

Selection

Selection InfoRationale

No valid atomic question selection. Terminate to generate answer directly...

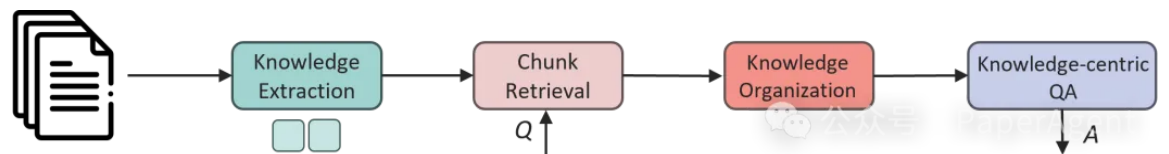
公众号 · PaperAgent

激活 Windows

例如，在患者历史病历搜索中，侧重于**事实信息检索能力**。主要挑战在于：

- 知识的理解和提取常常受到不恰当的知识切分的阻碍，破坏语义连贯性，导致检索过程复杂而低效；
- 常用的基于嵌入的知识检索受到嵌入模型对齐专业术语和别名的能力的限制，降低了系统准确率。

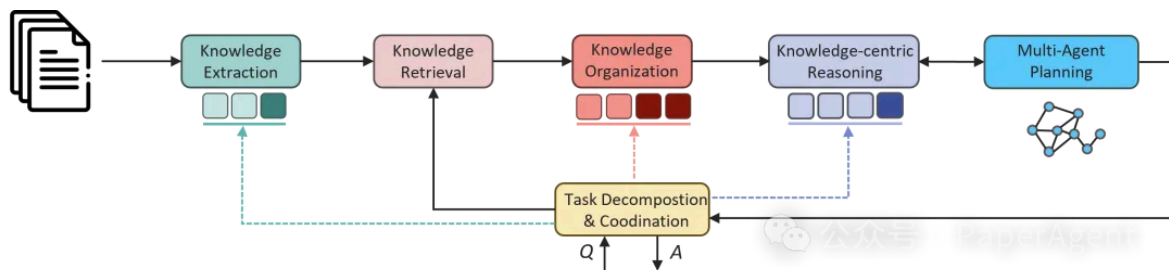
利用 PIKE-RAG，可以在知识提取过程中使用上下文感知切分技术、自动术语标签对齐技术和多粒度知识提取方法来提高知识提取和检索的准确率，从而增强事实信息检索能力，流程：



对于像为患者制定合理的治疗方案和应对措施建议这样的**复杂任务**，需要更高级的能力：

- 需要强大的领域特定知识才能准确理解任务并有时合理地分解任务；
- 还需要高级数据检索、处理和组织技术来预测潜在趋势；

而**多智能体规划**也将有助于兼顾创造力和可靠性。在这种情况下，可以初始化下面更丰富的管道来实现这一点。



与**Zero-Shot CoT**、**Naive RAG**、**Self-Ask**、**GraphRAG Local**、**GraphRAG Global**相比，PIKE-RAG 在准确率、F1 分数等指标上均表现出色，PIKE-RAG 在处理复杂推理任务方面具有显著优势，特别是在需要整合多源信息、进行多步骤推理的场景中。

Table 5: Performance comparison on 2WikiMultiHopQA. Best in bold, second-best underlined.

Method	EM	F1	Acc	Precision	Recall
Zero-Shot CoT	35.67	41.40	43.87	41.43	43.11
Naive RAG w/ R	51.20	59.74	62.80	59.06	62.30
Naive RAG w/ H-R	<u>51.40</u>	59.73	63.00	59.36	62.43
Self-Ask	23.80	37.49	51.60	34.56	60.72
Self-Ask w/ R	46.80	<u>64.17</u>	79.80	61.17	80.21
Self-Ask w/ H-R	48.00	63.99	<u>80.00</u>	<u>61.30</u>	<u>79.56</u>
GraphRAG Local	0.00	11.83	71.20	6.74	75.17
GraphRAG Global	0.00	7.35	45.00	4.09	55.43
Ours	66.80	75.19	82.00	74.04	78.87

Table 6: Performance comparison on MuSiQue. Best in bold, second-best underlined.

Method	EM	F1	Acc	Precision	Recall
Zero-Shot CoT	12.93	22.90	23.47	24.40	24.10
Naive RAG w/ R	<u>32.00</u>	43.31	44.40	<u>44.42</u>	47.29
Naive RAG w/ H-R	30.40	41.30	43.40	42.06	44.53
Self-Ask	16.40	27.27	35.40	26.33	37.65
Self-Ask w/ R	28.40	42.54	49.80	41.13	53.37
Self-Ask w/ H-R	29.80	<u>44.05</u>	<u>54.00</u>	42.47	<u>55.89</u>
GraphRAG Local	0.60	9.62	49.80	5.73	55.82
GraphRAG Global	0.00	5.16	44.60	2.82	52.19
Ours	46.40	56.62	59.60	57.45	59.53

首次提出了5级RAG系统能力与挑战，针对不同系统层级的技术挑战，PIKE-RAG框架都有针对性策略。

以下缩写被使用：“PA”代表文件解析，“KE”代表知识抽取，“RT”代表知识检索，“KO”代表知识组织，“KR”代表以知识为中心的推理。

Level	Challenges	Proposed Framework
L0	<ul style="list-style-type: none">Challenges arise in knowledge extraction due to the diverse formats of source documents, requiring sophisticated file parsing techniques.The construction of a high-quality knowledge base from raw, heterogeneous data introduces significant complexity in knowledge organization and integration.	<div>PA KE </div>
L1	<ul style="list-style-type: none">The understanding and extraction of knowledge are hindered by improper chunking, which disrupts semantic coherence, complicating accurate retrieval.Knowledge retrieval is impacted by the limitations of embedding models in aligning professional terminologies and aliases, reducing the system’s precision.	<div>PA KE RT KO KR</div>
L2	<ul style="list-style-type: none">Effective knowledge extraction and utilization are critical, as chunked text often contains both relevant and irrelevant information. Ensuring the retrieval of high-quality data is essential for accurate generation.The understanding and decomposition of tasks and rationale behind them often overlook the availability of supporting data, relying heavily on LLM capabilities.	<div>PA KE RT KO KR Task Decomp. & Coord.</div>
L3	<ul style="list-style-type: none">The challenges at this level center on knowledge collection and organization, which are vital for supporting predictive reasoning.LLMs have limitations in applying specialized reasoning logic, restricting their effectiveness in predictive tasks.	<div>PA KE RT KO KR Task Decomp. & Coord.</div>
L4	<ul style="list-style-type: none">The difficulty lies in extracting coherent logical rationales from complex knowledge bases, where interdependencies between multiple factors can result in non-unique solutions.The open-ended nature of creative questions complicates the evaluation of the reasoning and knowledge synthesis process, making it difficult to quantitatively assess answer quality.	<div>PA KE RT KO KR Multi-agent Plan. Task Decomp. & Coord.</div>

```
1 https://arxiv.org/abs/2501.11551
2 https://github.com/microsoft/PIKE-RAG
3 PIKE-RAG: sPecIalized KnowledgE and Rationale Augmented Generation
```

推荐阅读

- 动手设计AI Agents：Coze版（编排、记忆、插件、workflow、协作）
- [DeepSeek R1 + Agent 的下半场](#)
- RAG全景图：从RAG启蒙到高级RAG之36技，再到终章Agentic RAG！
- Agent到多模态Agent再到多模态Multi-Agents系统的发展与案例讲解（1.2万字，20+文献，27张图）

欢迎关注我的公众号“**PaperAgent**”，每天一篇大模型（LLM）文章来锻炼我们的思维，简单的例子，不简单的方法，提升自己。



PaperAgent

日更，解读AI前沿技术热点Paper

224篇原创内容

公众号

RAG专栏 87

RAG专栏 · 目录

上一篇 · 一篇最新大模型GraphRAG技术系统性综述

