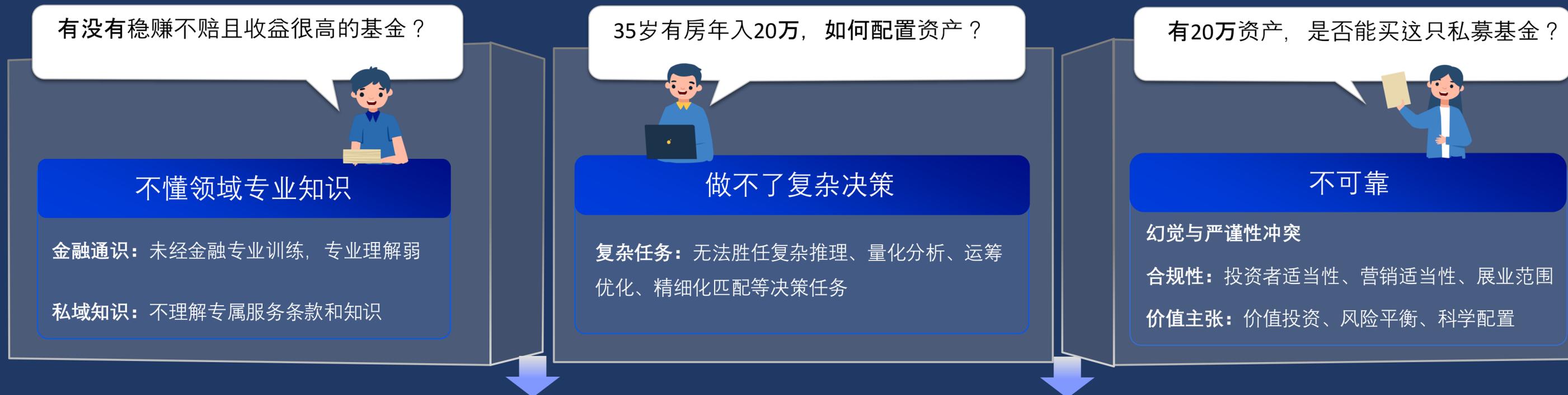


# OpenSPG-KAG框架及垂域应用

蚂蚁集团知识引擎负责人 梁磊

Github: <https://github.com/OpenSPG/KAG>

## 大模型实际产业落地遇到的“可信”挑战



- 生成式AI的技术特点，带来模型的创造性，但也带来极高的“bug rate”
- 生命攸关、财产攸关等敏感领域，如何确保可信、有效应用大模型服务产业，缺乏标准，技术成熟度差距较大
- 当下大模型技术创新关注在通识能力与模态扩展上，在特定领域知识强化、真实性、准确性提升上关注不够

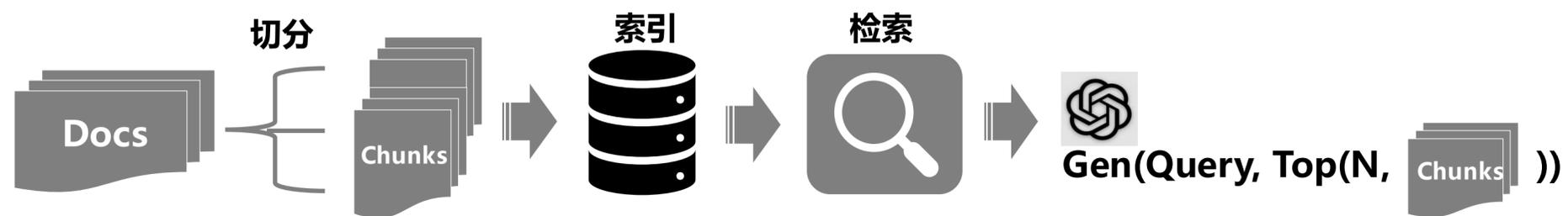
“可信”，是大模型大规模服务产业和社会的前提。不确保可信，就不会有产业变革与AGI时代

# 大纲

- 1、知识增强路线(检索、Graph、KG等)
- 2、垂域典型问题(逻辑、事实、语义等)
- 3、KAG框架设计(语义对齐、逻辑推理等)
- 4、KAG业务应用(医疗、政务、黑产、通用等)

# 知识增强大模型的主要路线

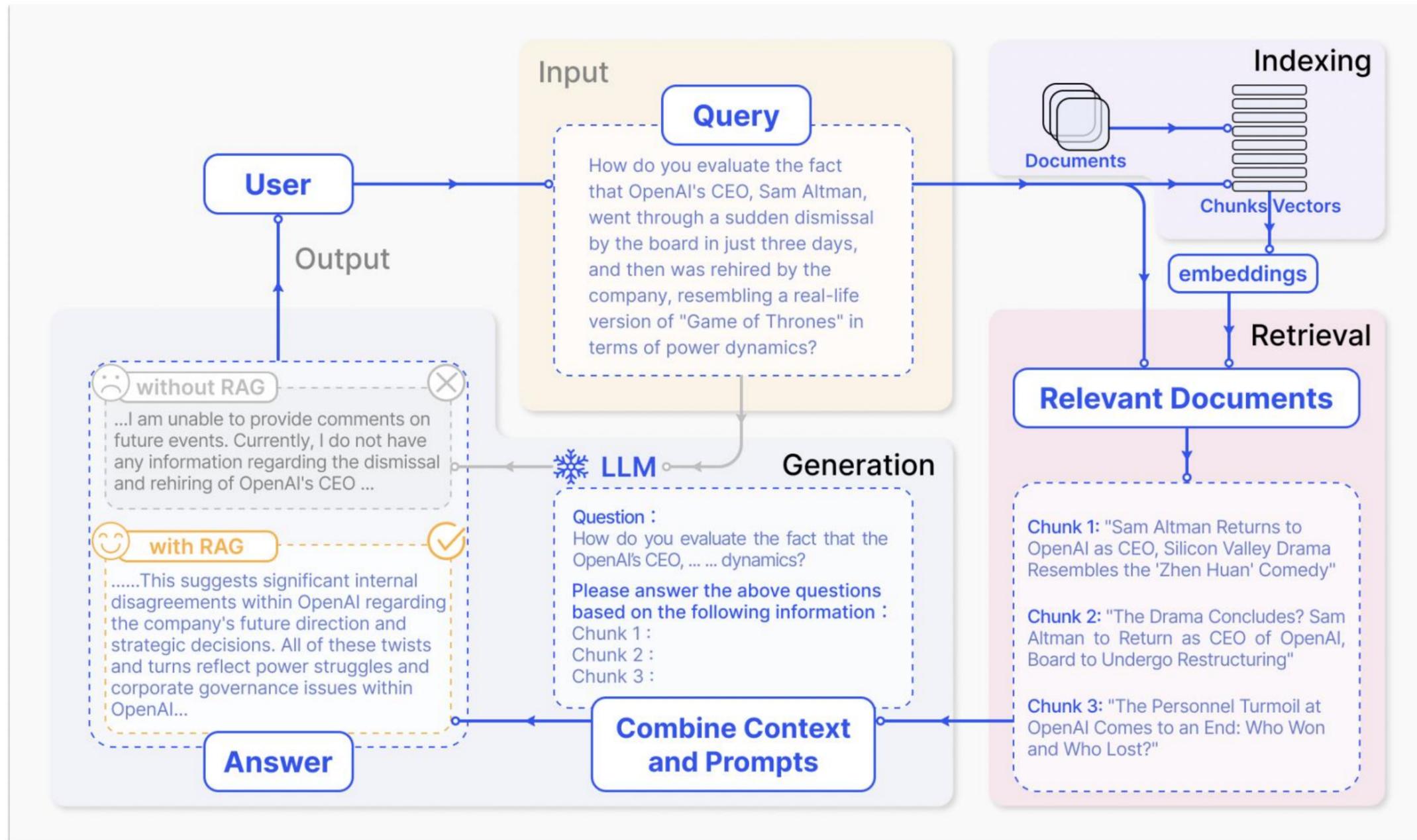
以文档检索为基础演进, Docs => R + G

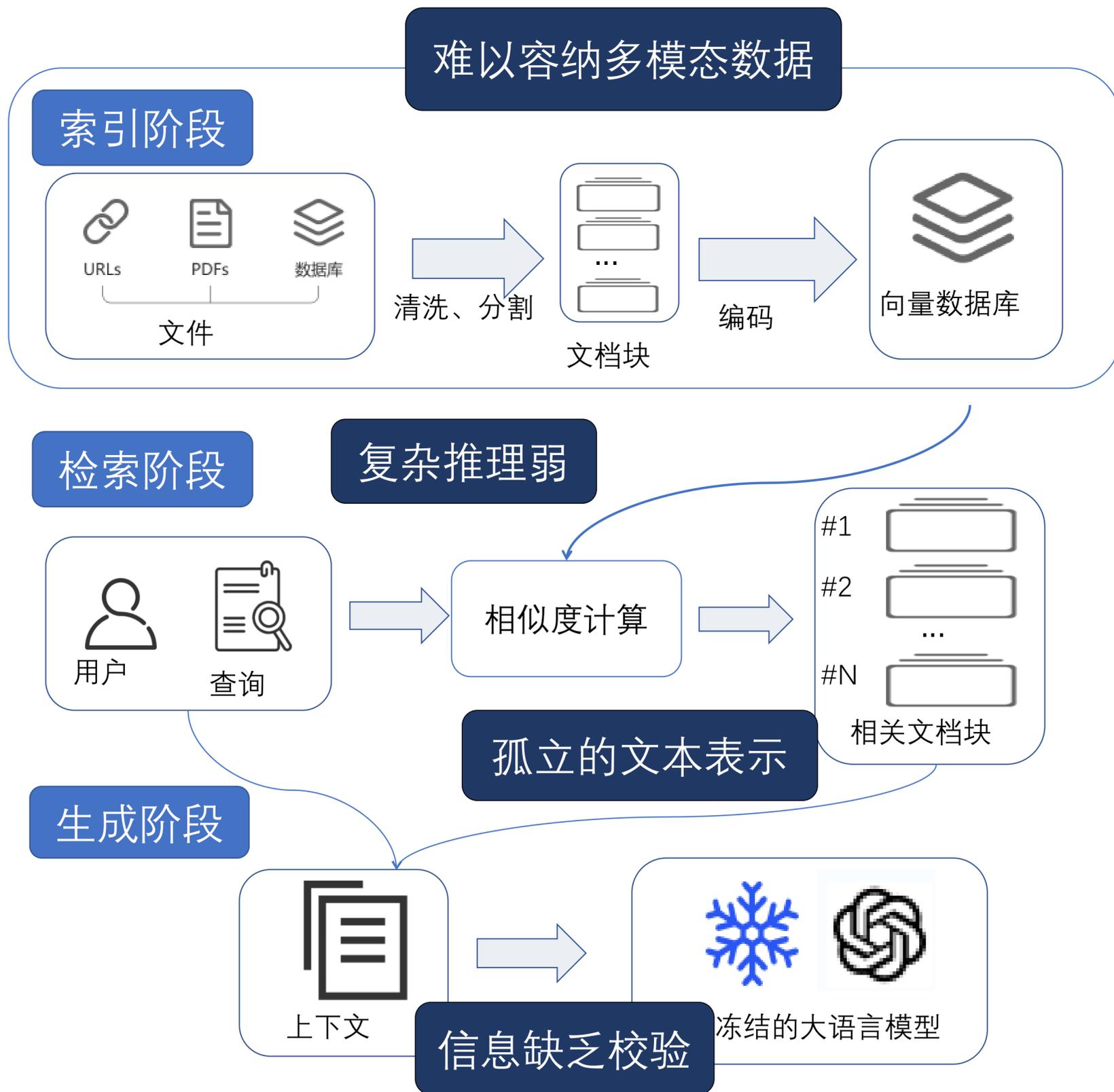


以检索(Docs Retrieval)为始、以LLM生成(Generation)为终

LLM 在回答问题或生成文本时，先会从大量文档中检索出相关的信息，然后基于这些信息来生成回答。

$$P(\text{query}, \text{Chunks} | \theta)$$





信息缺失，没有充分利用文档中的例如图像、表格等多模态数据

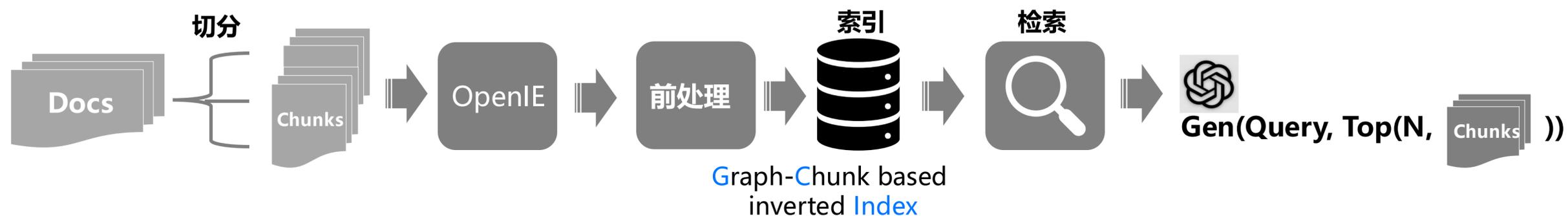
向量检索在面对复杂问题面临挑战，例如需要多跳、复杂逻辑。LLMs推理能力有限

依赖文本的语义相似度，只能确定文档之间的相似的，无法捕获不同之间的具体关系，以及为什么相关

检索到错误。噪声的信息后生成错误答案，仍然可能存在幻觉

# 知识增强大模型的主要路线

GraphRAG, KG增强文档索引, Docs => G-C Index => R + G



以检索(Docs Retrieval)为始、以LLM生成(Generation)为终



$$P(query, Summaries|\theta)$$

评估方式(主观评分):

- Comprehensiveness
- Diversity
- Empowerment

演进代表: LightRAG、DB-GPT等

Question: Which public figures are repeatedly mentioned across various entertainment articles?

**Naive RAG response**

Public figures who are repeatedly mentioned across various entertainment articles include <Public figures 1, 2, 3, and 4>. These individuals have been highlighted for various reasons, ranging from their professional achievements to their personal lives.

<Public figure 1> [...]  
<Public figure 2> [...]  
<Public figure 3> [...]  
<Public figure 4> [...]

These figures are frequently covered due to their high-profile status and the public's interest in their careers and personal lives. Their activities, whether in music, sports, or personal relationships, have significant cultural and economic impacts, as evidenced by the media coverage and public reactions.

**LLM evaluation**

**Comprehensiveness:** Winner = GraphRAG  
**Diversity:** Winner = GraphRAG  
**Empowerment:** Winner = GraphRAG

**GraphRAG response**

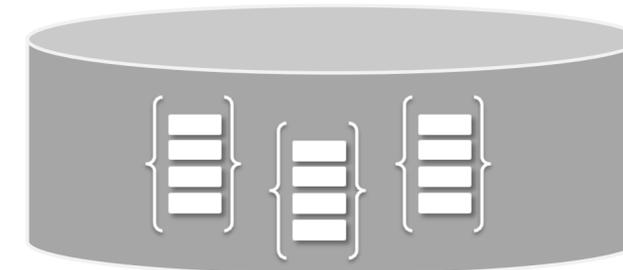
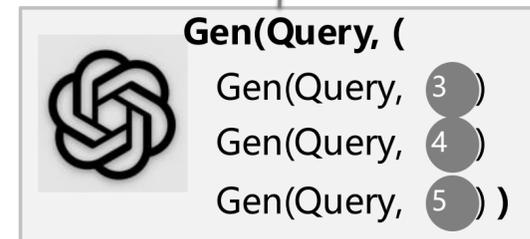
The entertainment industry is vast and diverse, encompassing film, television, music, sports, and digital media. Certain public figures stand out due to their significant contributions and influence across these sectors. The following summary highlights key individuals who are repeatedly mentioned in various entertainment articles, reflecting their impact and presence within the industry.

**Actors and Directors [... 20 figures ...]**  
**Public Figures in Controversy [... 3 figures ...]**  
**Musicians and Executives [... 5 figures ...]**  
**Athletes and Coaches [... 7 figures ...]**  
**Influencers and Entrepreneurs [... 3 figures ...]**

The repeated mention of these figures in entertainment articles signifies their ongoing relevance and the public's interest in their work. Their influence spans across various aspects of entertainment, from shaping cultural narratives in film and television to driving trends in music and digital media. These individuals not only contribute to their respective fields but also influence the broader cultural landscape, often becoming central figures in social discussions and public discourse.

代表作为《王右丞集》诗人，其家乡被司马迁称之为为什么？

王维是一位山水田园派诗人，以其代表作《王右丞集》而闻名，他的家乡被司马迁称之为“中华民族发祥地的核心区域”，这一称呼王维家乡的重要地位，以及其与中国历史和文化的紧密联系。王维的祖籍是... 总的来说，王维作为...



文档1

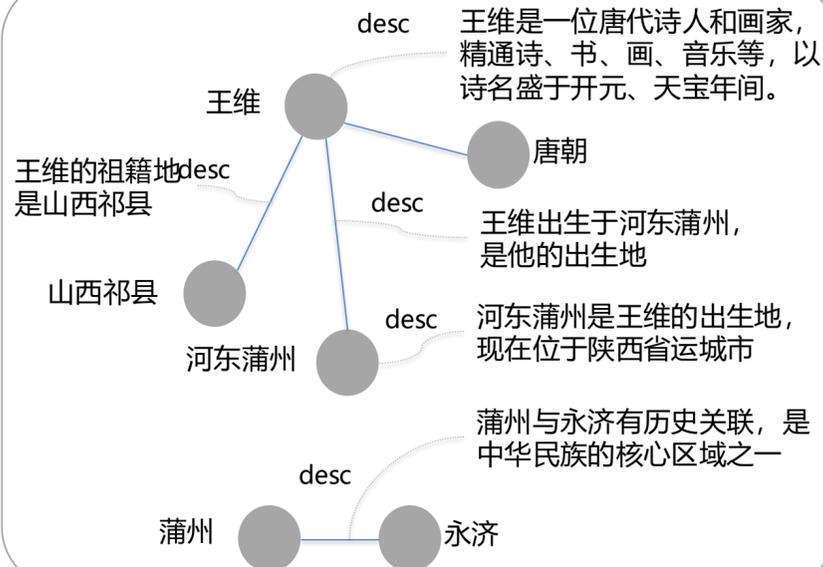
**王维**  
王维（693年或694年或701年—761年），字摩诘，号摩诘居士，河东蒲州人，祖籍山西祁县。唐朝诗人、画家。出身太原王氏分支河东王氏。于开元十九年（731年）状元及第，历官右拾遗、监察御史、河西节度使判官，...

文档2

**蒲州**  
古代地名。永济古称蒲坂，是中华民族发祥地的核心区域，司马迁在《史记》中称这里为“天下之中”。《路史》中记述：“柱所都蒲坂”，认为上古传说中的农神，即烈山氏之子柱曾建都于此，...

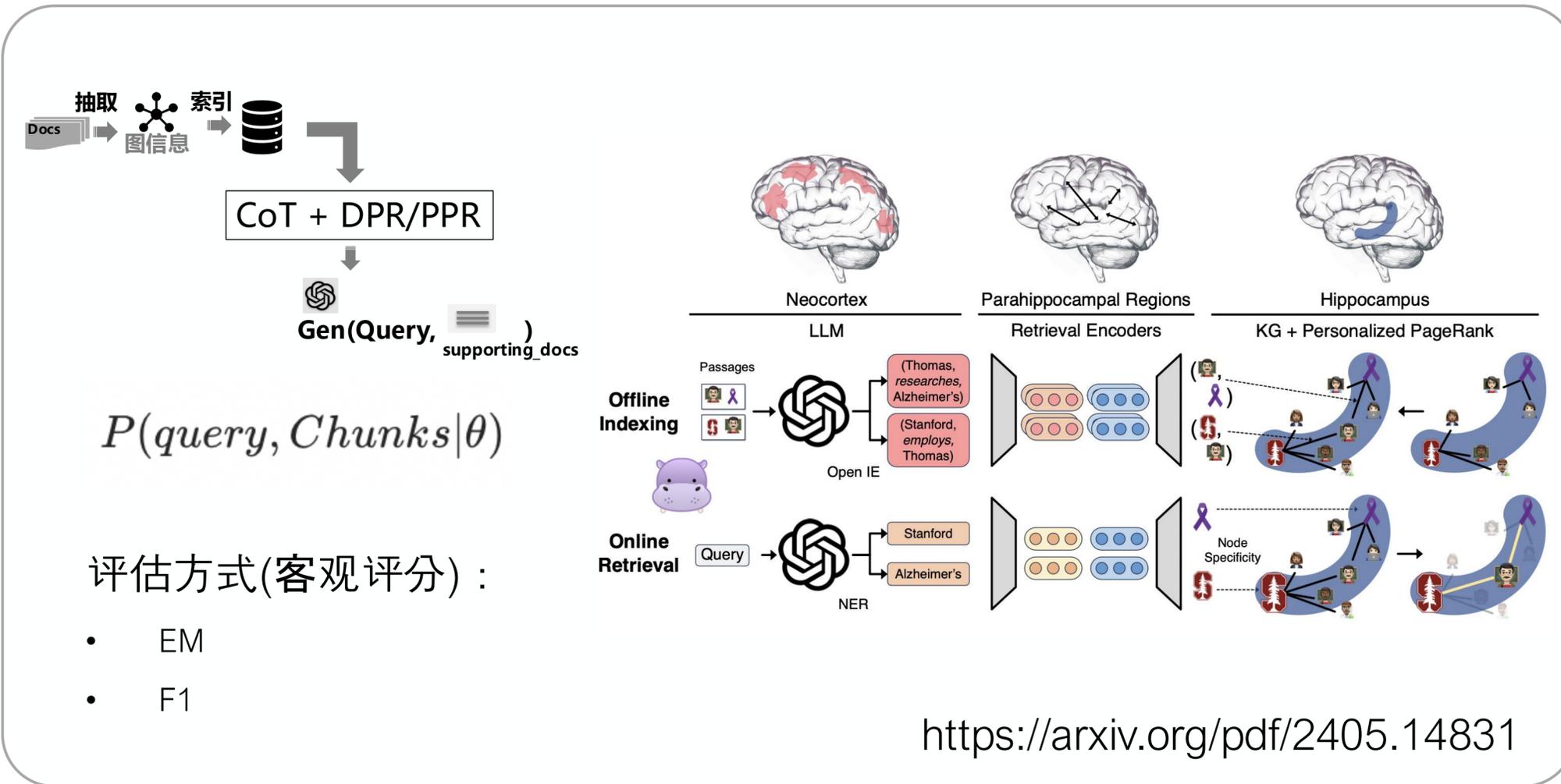


抽取



社区摘要

- 1 {"summary": "唐代作为历史背景", "explanation": "唐代是王维活动的历史背景，为他的艺术创作提供了背景。中国历史上的这一时期文化繁荣，政治稳定，是艺术发展的重要时期。王维与唐代的联系凸显了他当时文化景观中的作用。"}, {"summary": "山西祁县是王维的祖籍", "explanation": "山西祁县是王维的祖籍，这为了解他的家庭背景和文化遗产提供了见解。王维与山西祁县之间的联系揭示了他的根源和影响，塑造了他的艺术敏感性和创作成果。了解王维与山西祁县的联系可以丰富我们对个人历史和艺术动机的理解。"}, {"summary": "河东蒲州是王维的出生地", "explanation": "河东蒲州是王维的出生地，标志着这位有影响力的诗人和画家的地理起源。王维与河东蒲州的联系突显了他成长和早年生活经历的地域背景，这可能影响了他的艺术视野和主题选择。探究王维与河东蒲州的渊源，有助于我们了解他的成长经历和文化背景。"}, {"summary": "王维的主要成就和代表作", "explanation": "王维精通书画音乐，尤长五言诗，与孟浩然并称“王孟”，有“诗佛”之称，著有《相思》《山居秋暝》《王右丞集》..."}, {"summary": "永济是中华文明发祥地的历史见证", "explanation": "永济，古称蒲坂，是中华民族发祥地的核心区域，被司马迁在《史记》中称为“天下之中”。《路史》记载，“柱所都蒲坂”，指上古农神烈山氏之子柱曾建都于此。先秦史专家丁山考证认为，能植百谷百蔬的“柱”即为“后土”的省称..."}



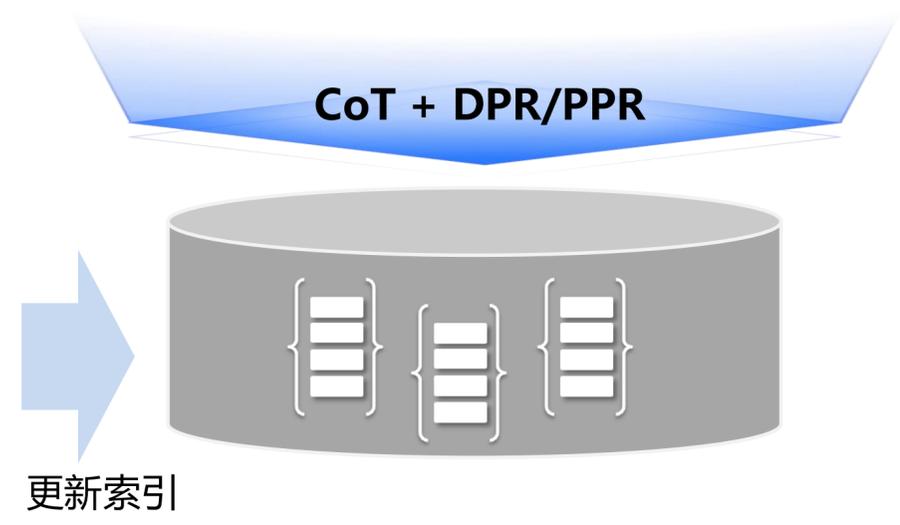
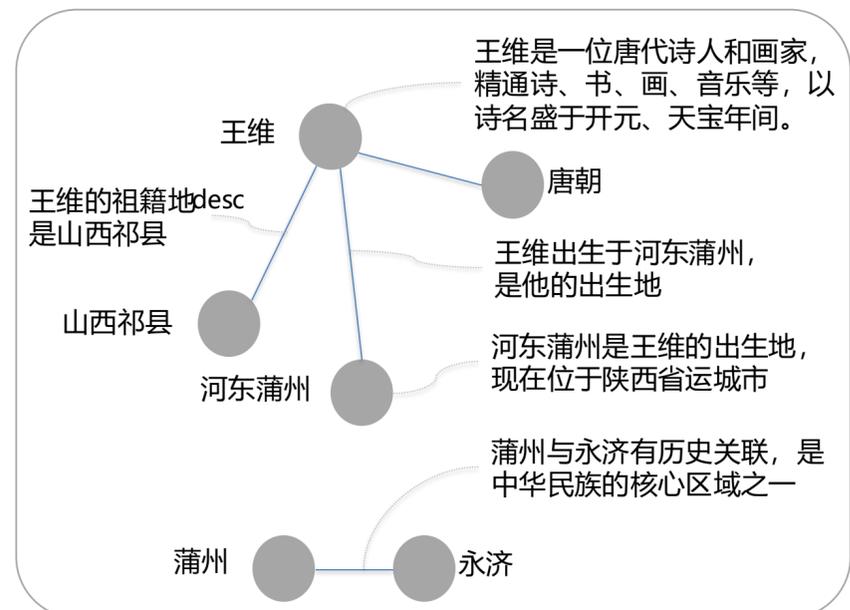
文档1

**王维**  
王维 (693年或694年或701年—761年)，字摩诘，号摩诘居士，河东蒲州人，祖籍山西祁县。唐朝诗人、画家。出身太原王氏分支河东王氏。于开元十九年 (731年) 状元及第，历官右拾遗、监察御史、河西节度使判官，....

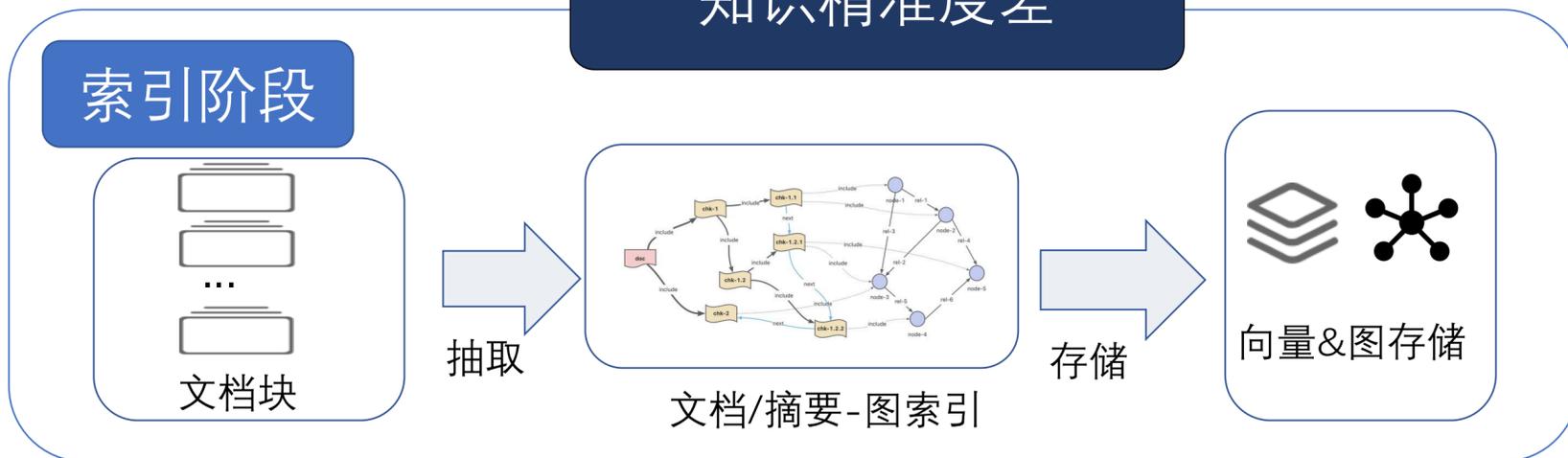
文档2

**蒲州**  
古代地名。永济古称蒲坂，是中华民族发祥地的核心区域，司马迁在《史记》中称这里为“天下之中”。《路史》中记述：“柱所都蒲坂”，认为上古传说中的农神，即烈山氏之子柱曾建都于此，....

抽取

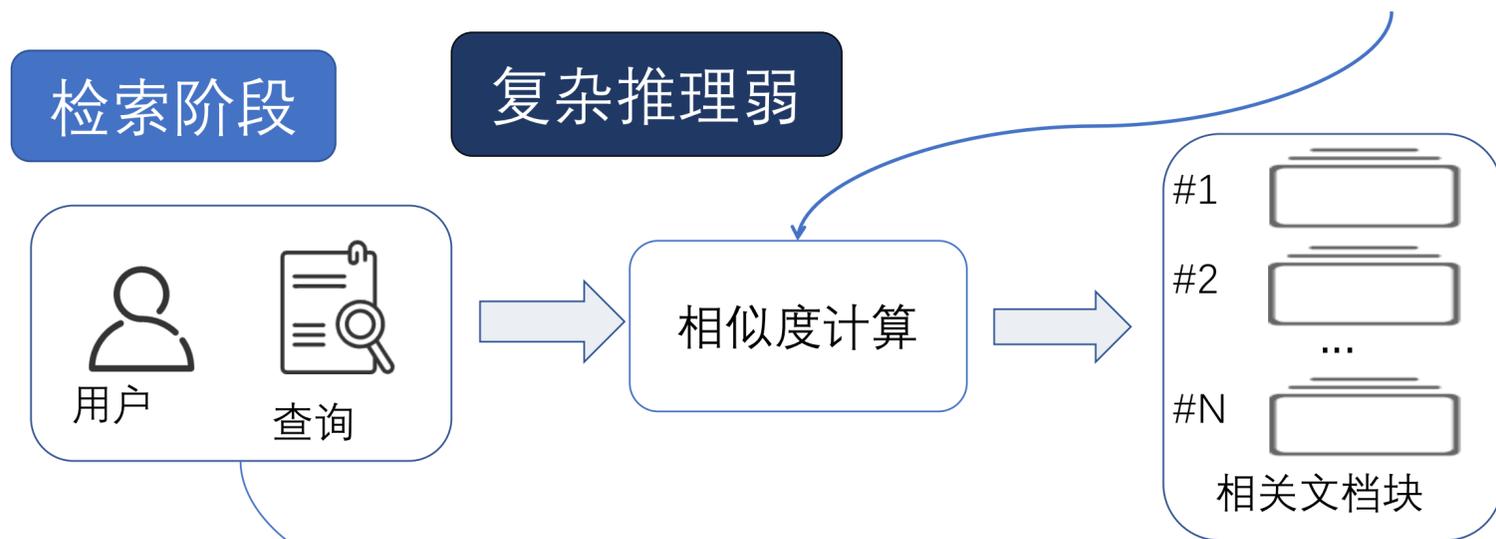


## 知识精准度差



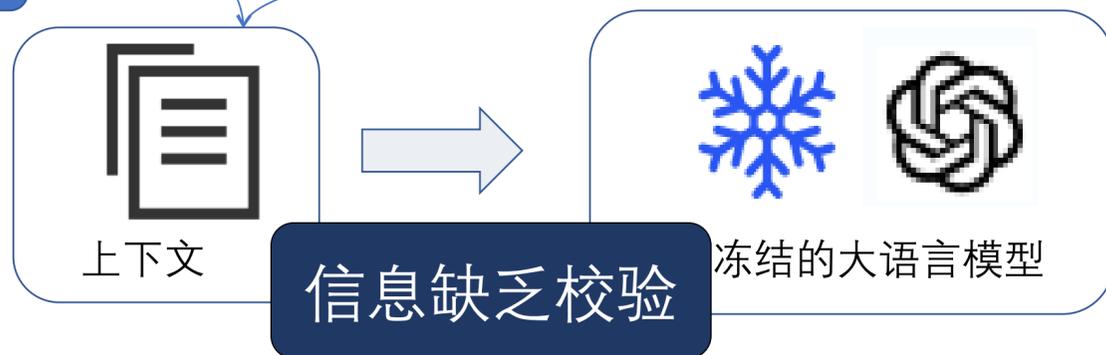
依赖开放信息抽取方法，引入了大量的数据噪声，知识精准度差。

## 复杂推理弱



使用图结构实现跨文档的信息传播，缓解了向量召回的不足。例如需要多跳、复杂逻辑依然存在不足。LLMs推理能力有限

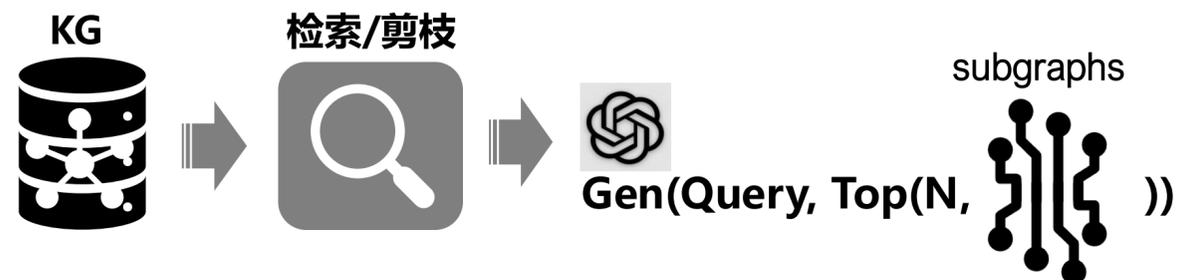
## 生成阶段



检索到错误。噪声的信息后生成错误答案，仍然可能存在幻觉

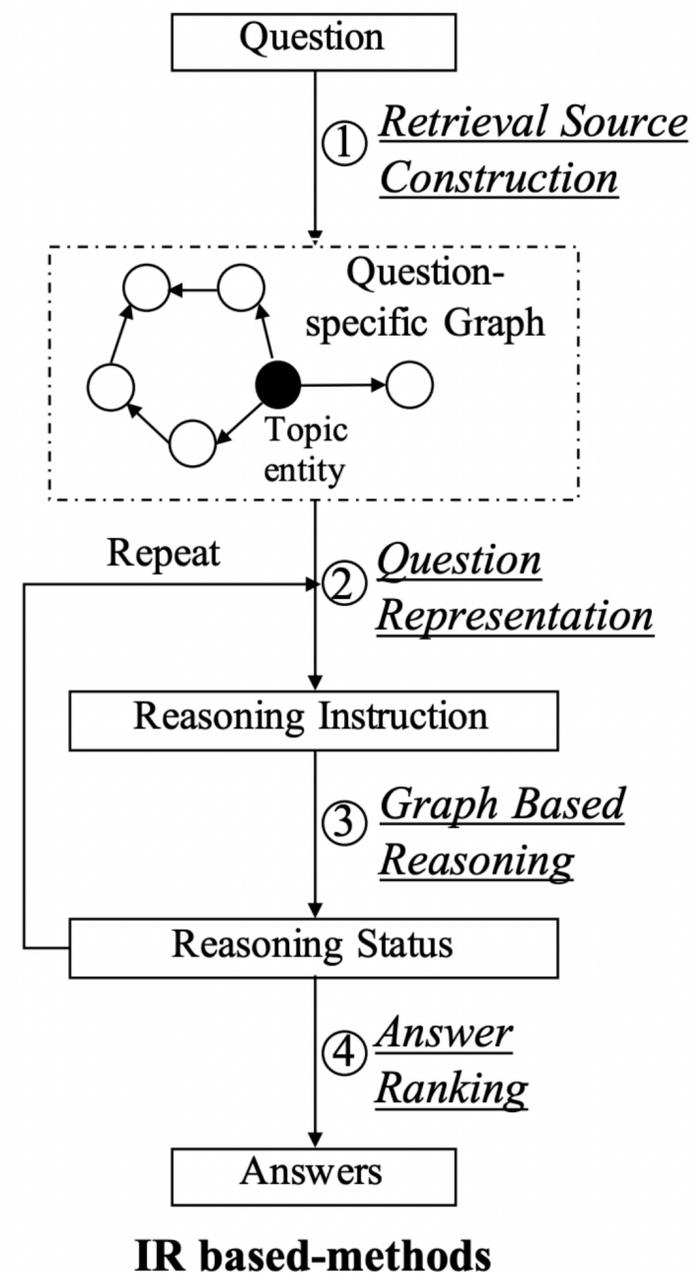
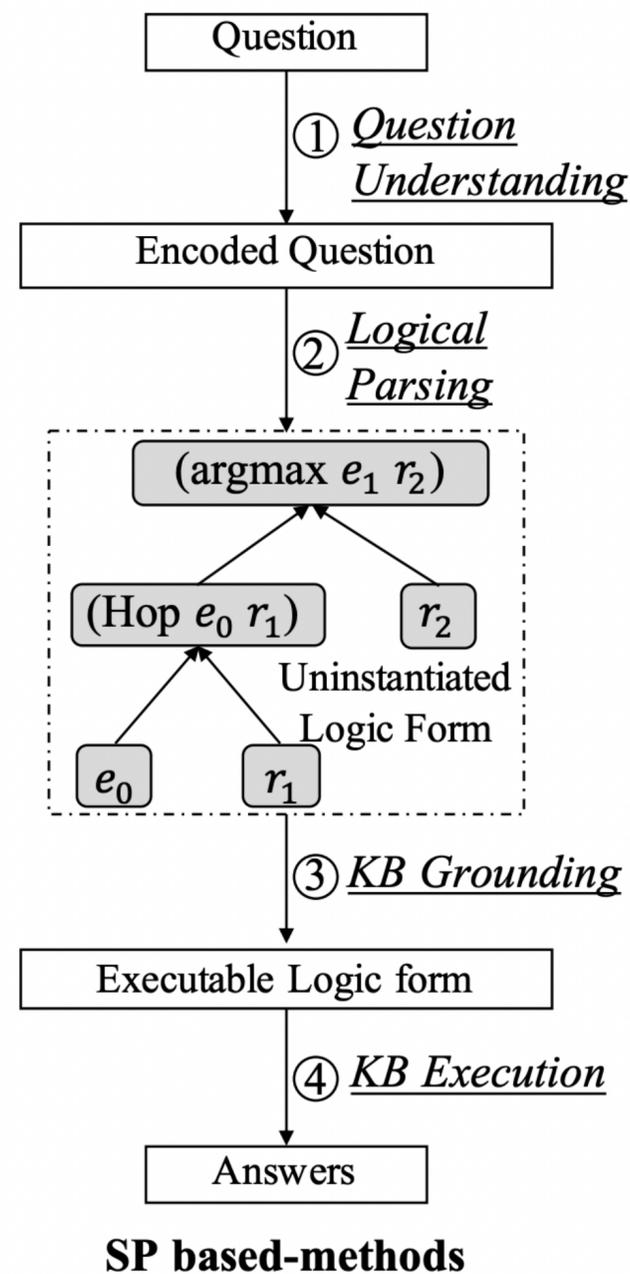
# 知识增强大模型的主要路线

以KBQA为原型演进, KGs => R + G

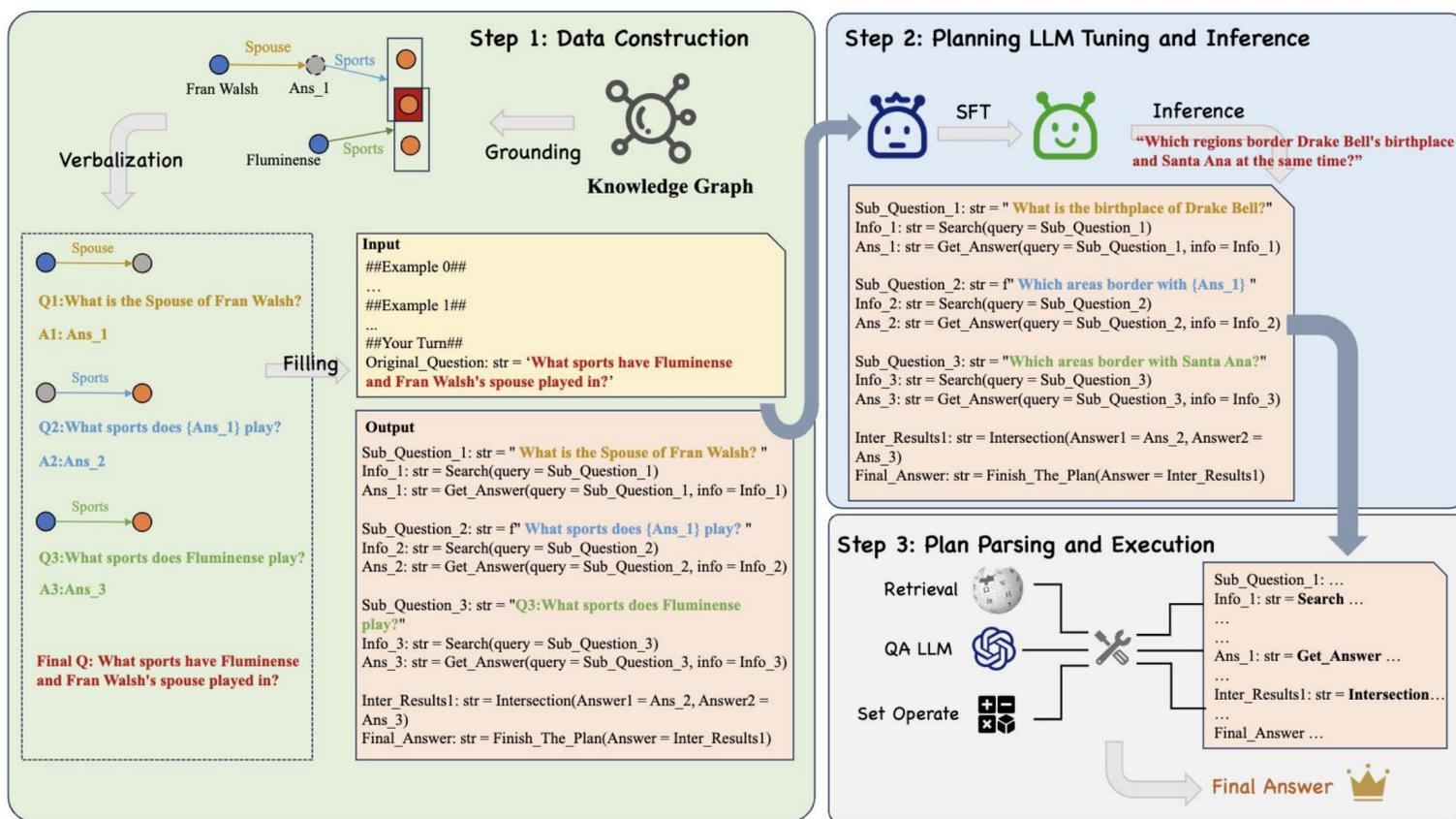


以检索(Graph Retrieval)为始、以LLM生成(Generation)为终

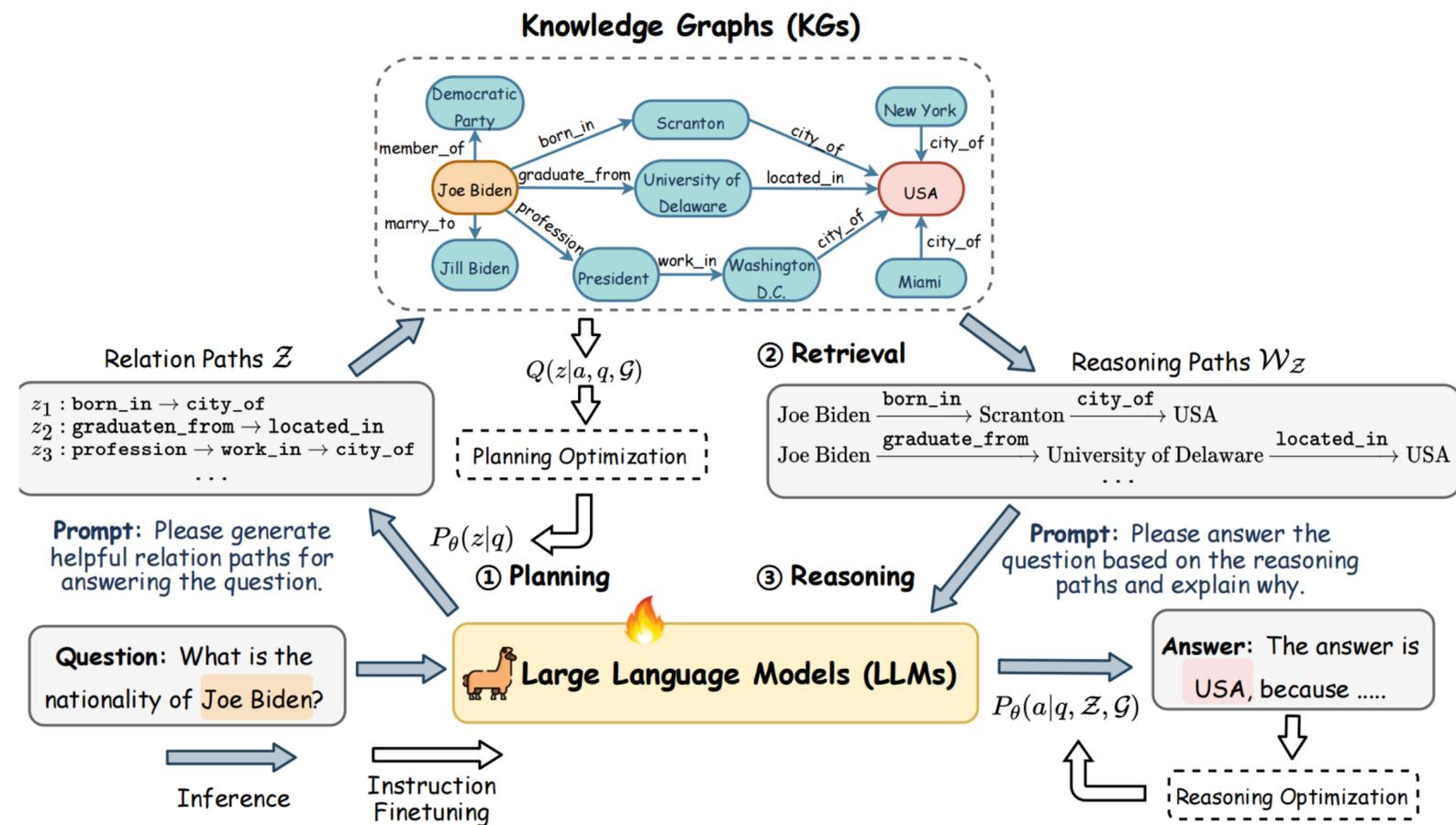
问题理解  
↓  
逻辑拆解  
↓  
知识查询  
↓  
逻辑执行



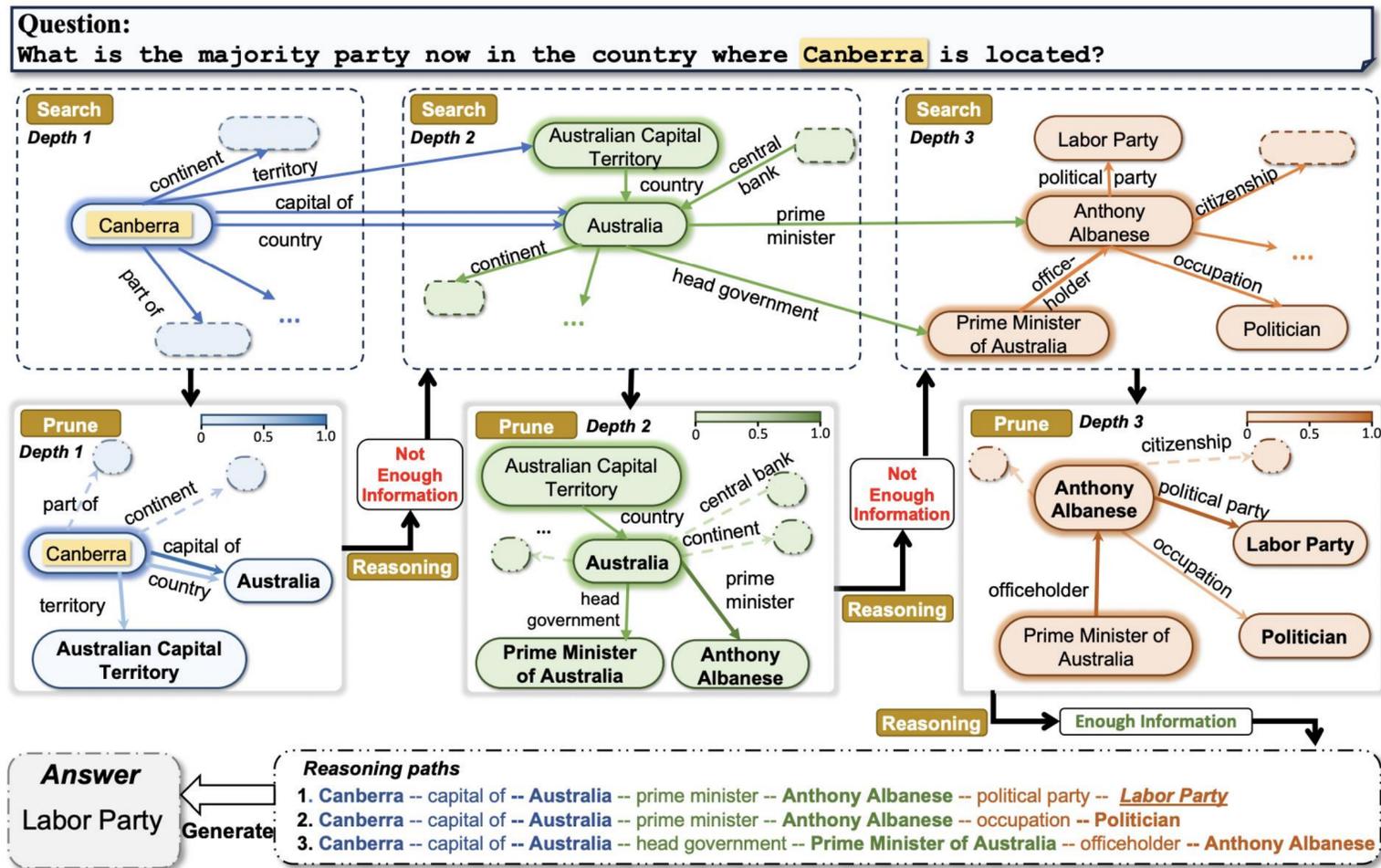
子图检索  
↓  
相关度量  
↓  
剪枝推理  
↓  
结果排序



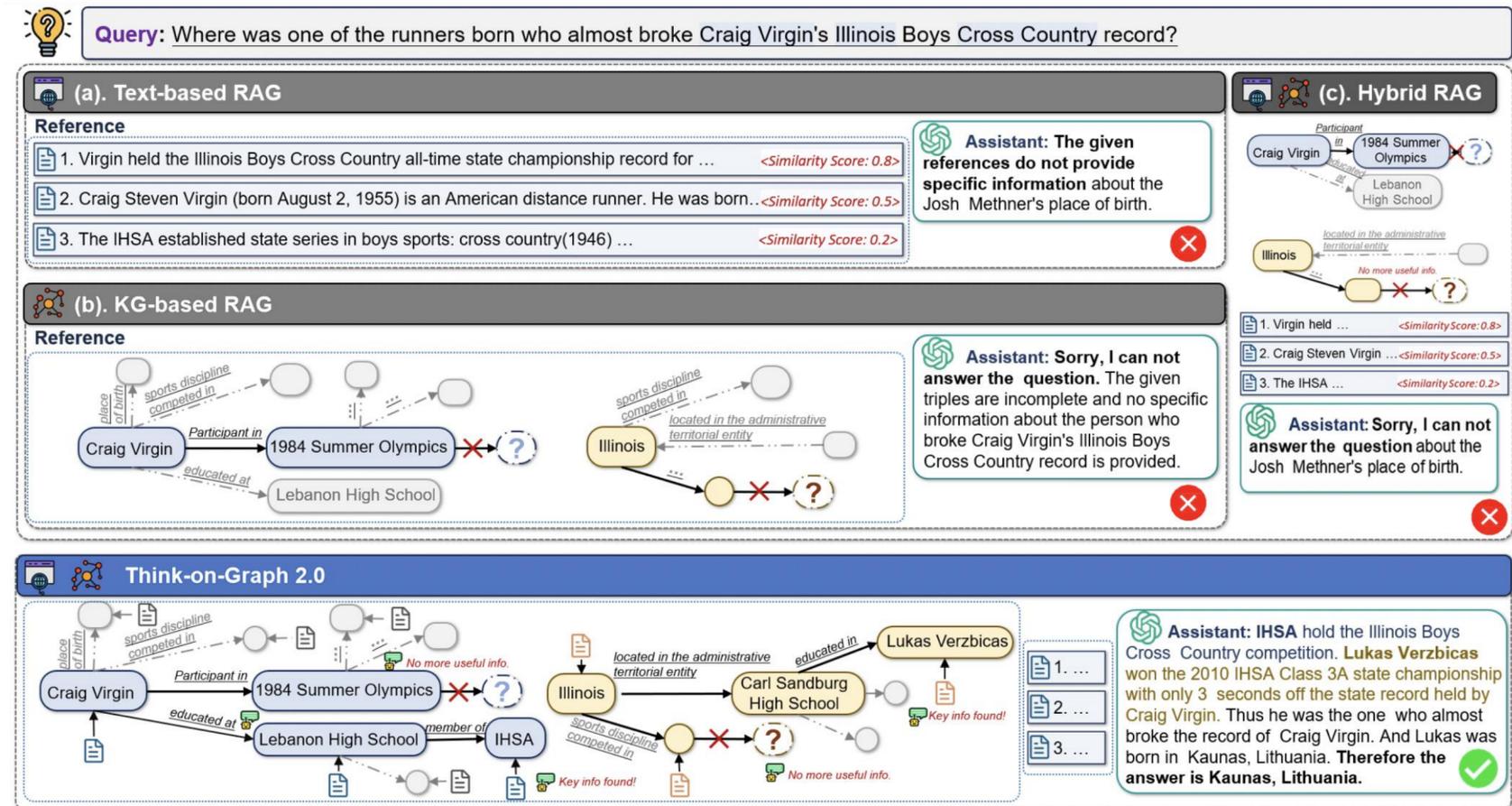
Learning to Plan for Retrieval-Augmented Large Language Models from Knowledge Graphs



Reasoning on Graphs: Faithful and Interpretable Large Language Model Reasoning



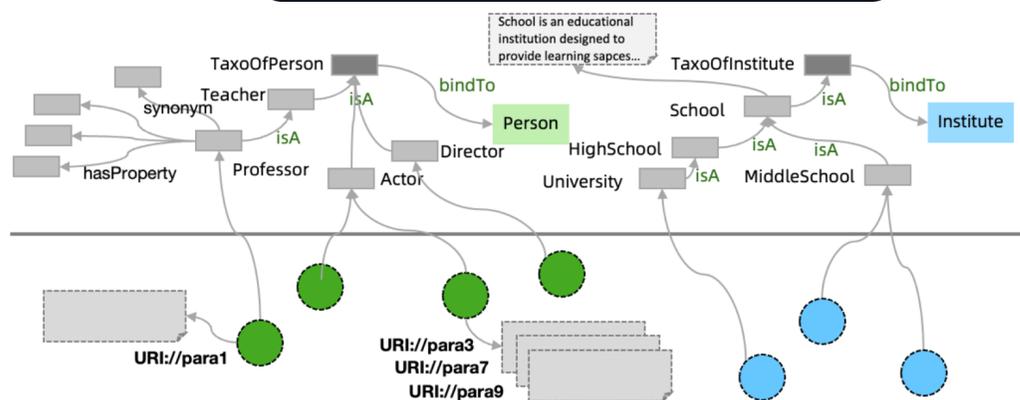
THINK-ON-GRAPH: DEEP AND RESPONSIBLE REASONING OF LARGE LANGUAGE MODEL ON KNOWLEDGE GRAPH



THINK-ON-GRAPH 2.0: DEEP AND FAITHFUL LARGE LANGUAGE MODEL REASONING WITH KNOWLEDGE GUIDED RETRIEVAL AUGMENTED GENERATION

## 图谱构建

### 图谱构建门槛高



依赖高质量的知识图谱，图谱构建门槛高，高质量的知识图谱需要大量人力

## 检索阶段

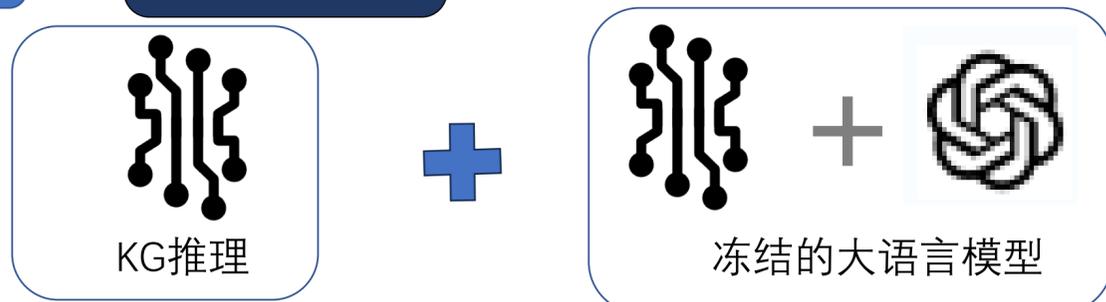
### 实体粒度偏差



信息损失大，知识图谱中只包含了实体、关系、属性等，相对信息丰富的原始文本，信息损失较大

## 生成阶段

### 可读性差



可读性差，生成的答案包含关键事实，上下文信息较少

**垂直领域的知识问答，存在哪些问题？**

福建龙海农村商业银行股份有限公司在公司治理方面，近期遭遇了监管处罚。

[T][1]2023年7月，该银行因未按规定进行客户身份识别和大额交易报告，被中国人民银行漳州市中心支行罚款66万元，这反映出其在合规管理上存在一定的问题。

Yu' n' fei  
非巨额罚款

在资产质量方面，截至2023年底，福建龙海农村商业银行的不良贷款率为0.98%，关注类贷款率为4.7%，拨备覆盖率为571.08%。[T][1]这些数据显示，尽管信贷资产质量相对稳定，但拨备覆盖水平略低，可能对未来的风险抵补能力构成一定压力。

Yu' n' fei  
判断不准确

在资产质量方面，截至2023年底，廊坊银行的不良贷款率为2.03%，关注类贷款率为3.91%，这表明其信贷资产质量存在一定的压力。[T][1]尽管如此，拨备覆盖率为153.19%，意味着银行有一定的风险抵补能力。

kjkiko  
这个拨备水平.....还有风险抵补能力

利润为2.9亿元，总资产收益率为1.19%，净资产收益率为11.23%。[T][1]然而，净息差从上一年度下降了0.38个百分点至2.65%，表明其盈利能力受到一定影响。

Yu' n' fei  
2.65%的净息差，盈利能力还是很强

[T]

在评级调整方面，联合见智于2022年7月6日将廊坊银行的评级从BBBf上调至BBBf+，显示出评级机构对其信用状况的改善持乐观态度。[T][Ra][3]

kjkiko  
虽然是上调几倍，但都BBB了，还乐观.....

在资产质量方面，河北银行的不良贷款率为1.31%，关注类贷款率为2.76%，拨备覆盖率为175.04%，虽然拨备覆盖水平尚可，但信贷资产质量的恶化趋势值得关注。[T][2][R]

kjkiko  
因计提信用减值导致增收不利，资产质量本身并未恶化

【65.82】分，在全省地级行政区排名 1/11，在全国排名 66/416，属于【中低】风险地区。该区域评价摘要：**赣州市经济实力在江西省排名首位**，第二产业对经济贡献较大，第三产业对经济增长贡献不大。赣州市财政自给能力较弱，综合财力依赖土地出让收入。赣州市地方政府债务负担可控，广义债务负担不重；区域投融资平台有活力，未来一年到期城投债券占存量城投债券比较均衡。←

jing' lue  
与事实不符←

赣州银行股份有限公司在 2023 年的运营状况显示了其在多个关键领域的挑战和机遇。**[T][1]**在公司治理方面，2023 年 11 月 27 日，**赣州银行的 8000 万股江西银行股权拍卖流拍**，这可能影响其股权结构的稳定性。**[T][2]**同时，核准王舒君的董事任职资格，表明管理层正在经历调整，可能影响决策效率和执行能力。**[4]**←

jing' lue  
无事实依据←

在资产质量方面，信贷资产质量成为关注点，不良贷款率为 1.35%，关注类贷款率为 0.45%，显示出信贷结构风险。**[T][R]**同时，**投资踩雷的风险提示可能意味着投资组合中存在潜在损失。**←

Guo Yanfang  
应该没有投资踩雷←

四川银行股份有限公司的现状分析如下：←

jing' lue  
无事实依据←

在公司治理方面，虽然**注册资本金在全国城商行中居首**，但存在一定的监管处罚记录。**[1]**2022 年 5 月 9 日，银行因办理房地产开发贷款业务违规被罚款 40 万元，显示了在合规管理上存在一定的问题。**[3]**此外，2022 年 11 月，攀钢系股东出售了部分股权，占总股本的 0.3225%，这表明股东结构有所变动。**[4]**←

jing' lue  
数据滞后←

## 向量相似召回

```
⊖{  
  "query": "怎么查我的养老金",  
  "label": "参保人员参保信息查询",  
  "sim_dict": "{\"养老保险接续申请\": \"0.6829\", \"企业离退休人员重复领取养老保险待遇暂停养老金发放\": \"0.6581\", \"退休高级职称人员增加养老金待遇资格确认\": \"0.6055\"}"  
}
```



```
⊖{  
  "query": "有没有咨询电话能问问身份证丢了重新申领的进度?",  
  "sim_dict": "{\"居民身份证首次申领|咨询电话\": \"0.6562\", \"户籍事项证明|咨询电话\": \"0.6475\", \"居民身份证挂失申报|咨询电话\": \"0.636\"}"  
}
```

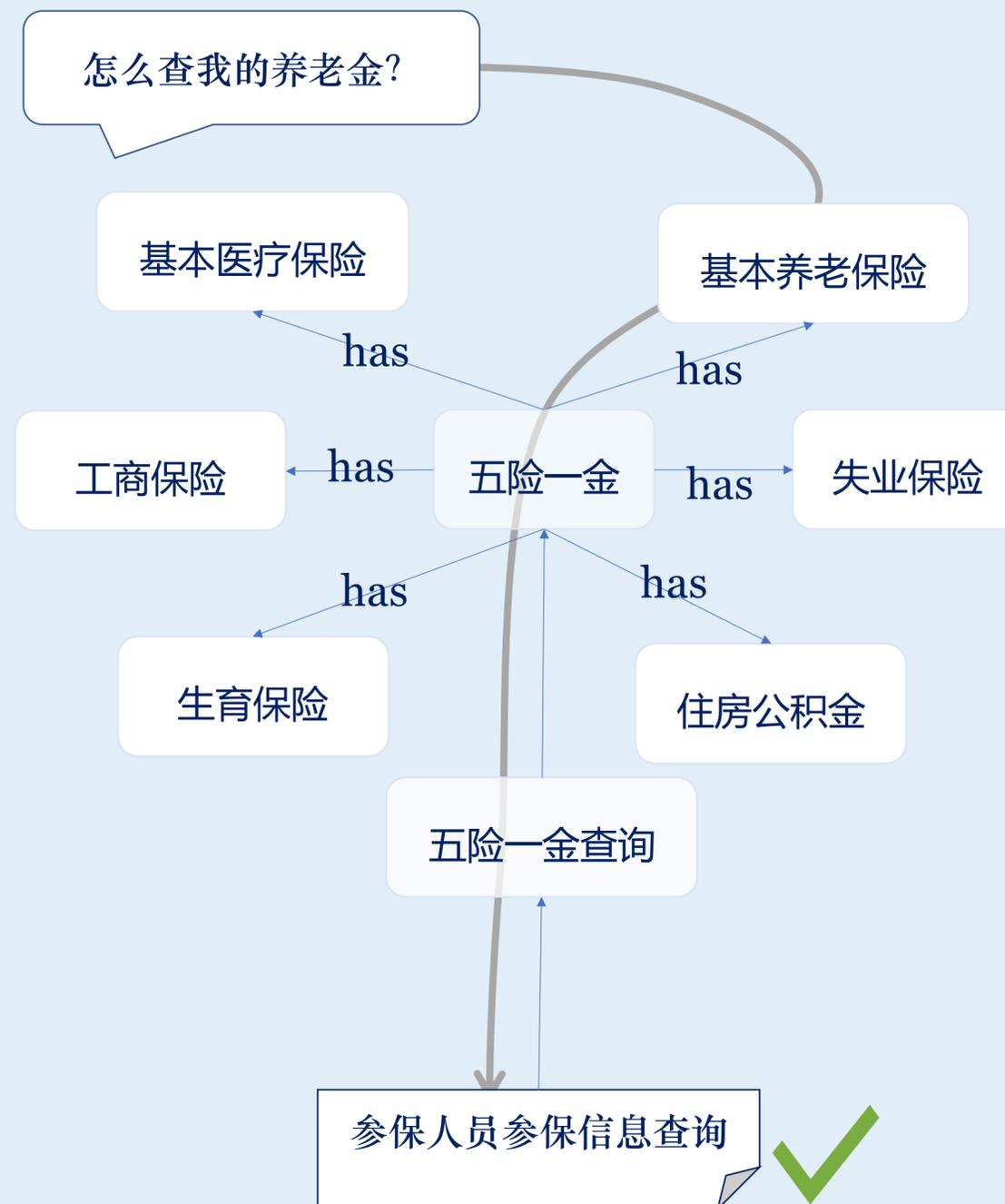


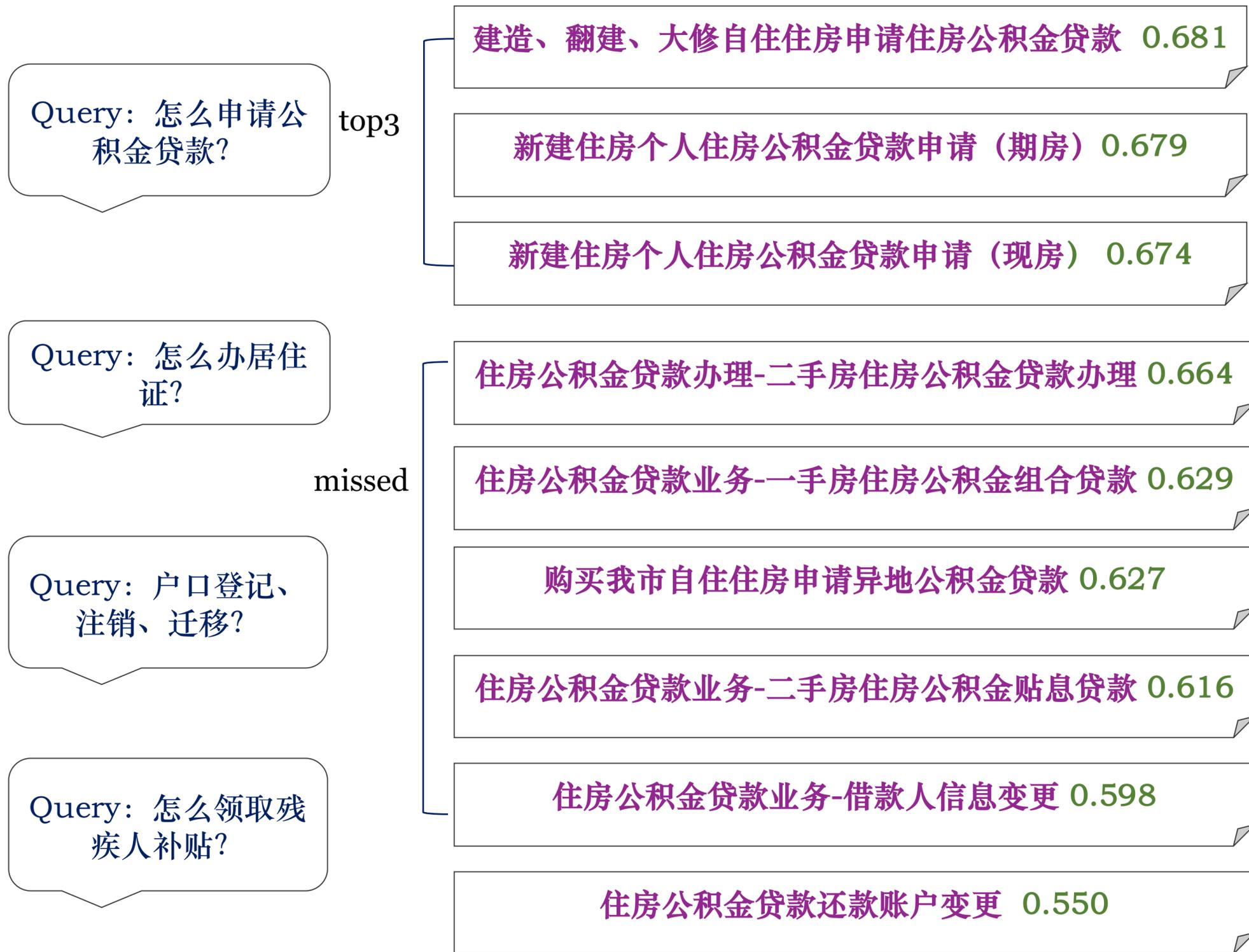
```
⊖{  
  "query": "就业多久可以办理合法稳定就业居住证",  
  "label": "合法稳定就业居住证办理|受理条件",  
  "sim_dict": "{\"合法稳定就业居住证办理|工作时间\": \"0.8057\"}"  
}
```



语义上可以推断是问“条件”，向量相似更接近“时间”

## 查询逻辑







1. 错误定性或错误逻辑

2. 事实性错误或无依据

3. 时间、数值不敏感

4. 张冠李戴

5. 不能区分重要性

6. 语义不精准

7. 召回不完备



**Indexing**

**Retrieve**

**Generation**

**优点**

**缺点**

**RAG**

-->

Chunk&  
向量化

向量检索

QFS

构建成本低  
效率高  
信息完备

精准性不足  
逻辑性差

**GraphRAG**

-->

Chunk&  
向量化&  
图

向量检索  
&PPR

QFS

使用门槛低  
文档语义关联  
信息完备

精准性不足  
逻辑性差

**KBQA**

-->

实体抽取  
SPO抽取

结构化查询语句  
实体链指  
语义推理

引用原事实

精准  
逻辑清晰  
置信度高  
时间和数值敏感

构建成本高  
信息有损  
知识缺失  
可读性差

## Doc Retrieval-Based RAG依赖LLM的结果生成

矛盾  
错误

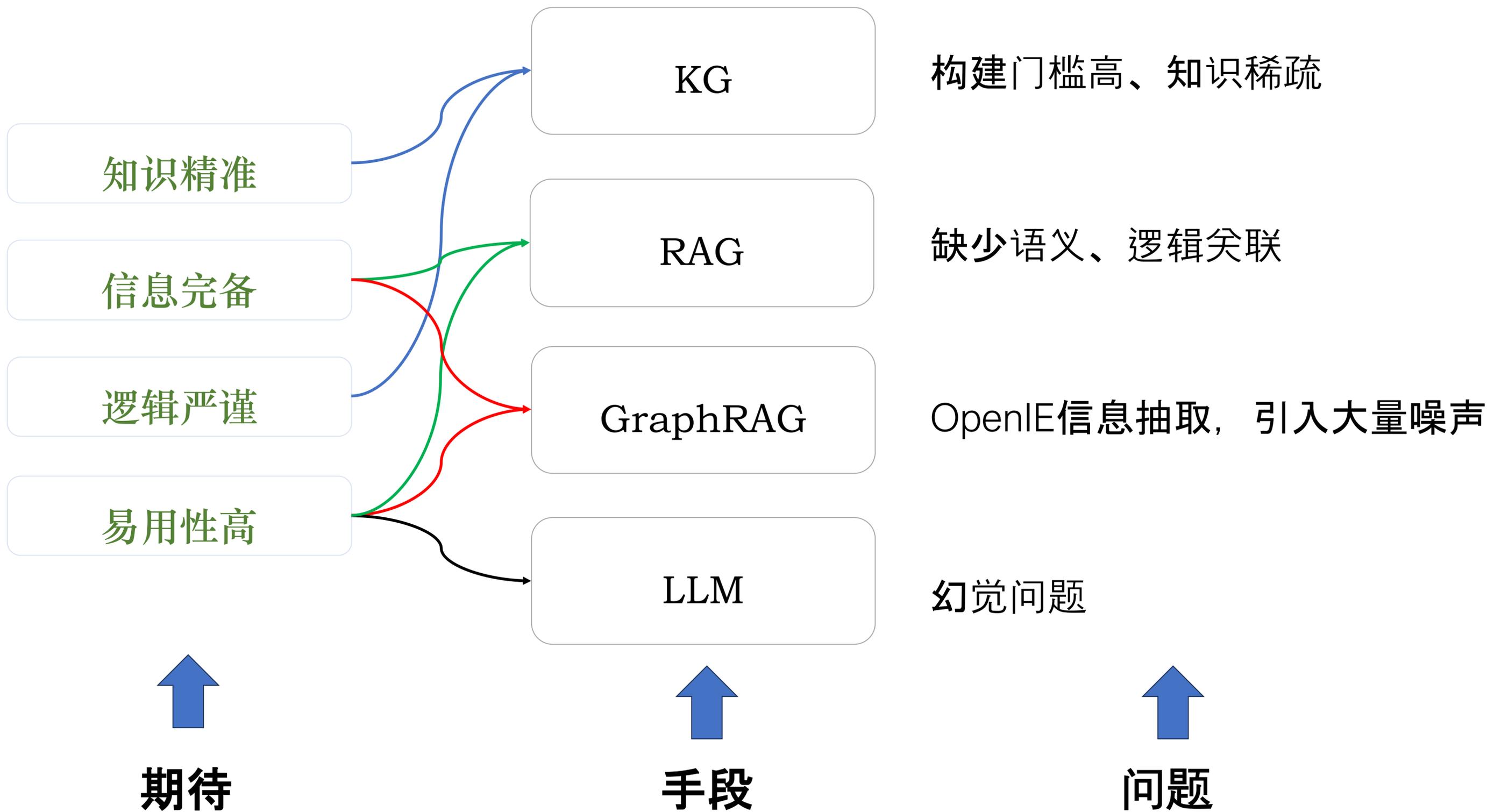
实体  
反转

合并  
错误

概念  
替换

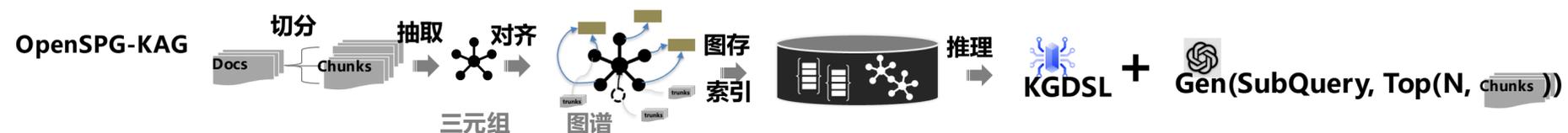
Error Type	Original Text	Factual Inconsistent Text
KCont.	功能饮料中的维生素、矿物质等，对于运动后快速补充身体营养， <b>消除</b> 疲劳具有一定作用。 The vitamins and minerals in energy drinks play a certain role in quickly replenishing nutrients and <b>eliminating</b> fatigue after exercise.	功能饮料中的元素、微生物等，对于运动后快速补充身体营养， <b>增加</b> 疲劳具有一定作用。 The vitamins and minerals in energy drinks play a certain role in quickly replenishing nutrients and <b>inducing</b> fatigue after exercise.
KInve.	一般蚕可以活一个多月，其中从孵化到结茧根据季节不同大约是 <b>25-32</b> 天，变成蛹后有 <b>15-18</b> 天，最后成蛾是1-3天。 A typical silkworm can live for just over a month, during which the period from hatching to cocooning varies roughly from <b>25 to 32</b> days depending on the season, followed by <b>15 to 18</b> days as a pupa, and finally 1 to 3 days as a moth.	一般蚕可以活一个多月，其中从孵化到结茧根据季节不同大约是 <b>15-18</b> 天，变成蛹后有 <b>25-32</b> 天，最后成蛾是1-3天。 A typical silkworm can live for just over a month, during which the period from hatching to cocooning varies roughly from <b>15 to 18</b> days depending on the season, followed by <b>25 to 32</b> days as a pupa, and finally 1 to 3 days as a moth.
KConf.	防晒霜中的 <b>无机</b> 化学物质可以 <b>反射或散射</b> 皮肤上的光线，而 <b>有机(碳基)</b> 化学物质可以吸收紫外线。 The <b>inorganic chemicals</b> in sunscreen can <b>reflect or scatter light on the skin</b> , while <b>organic (carbon-based) chemicals</b> can <b>absorb ultraviolet rays</b> .	防晒霜中的 <b>无机</b> 化学物质和 <b>有机(碳基)</b> 化学物质都可以 <b>反射或散射</b> 皮肤上的光线、 <b>吸收</b> 紫外线。 Both the <b>inorganic chemicals</b> and <b>organic (carbon-based) chemicals</b> in sunscreen can <b>reflect or scatter light on the skin</b> and <b>absorb ultraviolet rays</b> .
KConc.	随着健康意识的增强，越来越多的人开始注重 <b>膳食平衡</b> 。 With the increasing awareness of health, more and more people are beginning to focus on <b>a balanced diet</b> .	随着健康意识的增强，越来越多的人开始注重 <b>膳食的有机质量</b> 。 With the increasing awareness of health, more and more people are beginning to focus on <b>the organic quality of their diets</b> .

- 蚂蚁公布了**FCE(factual consistency evaluation) benchmark**，定义了RAG情况下会出现的“幻觉”问题
- 公开模型，都存在不同种类的幻觉偏差，**30%~40%**的RAG会存在结构性错误，不易被察觉。
- 除了文本的检索增强，我们需要**更加知识化的表达**，降低幻觉



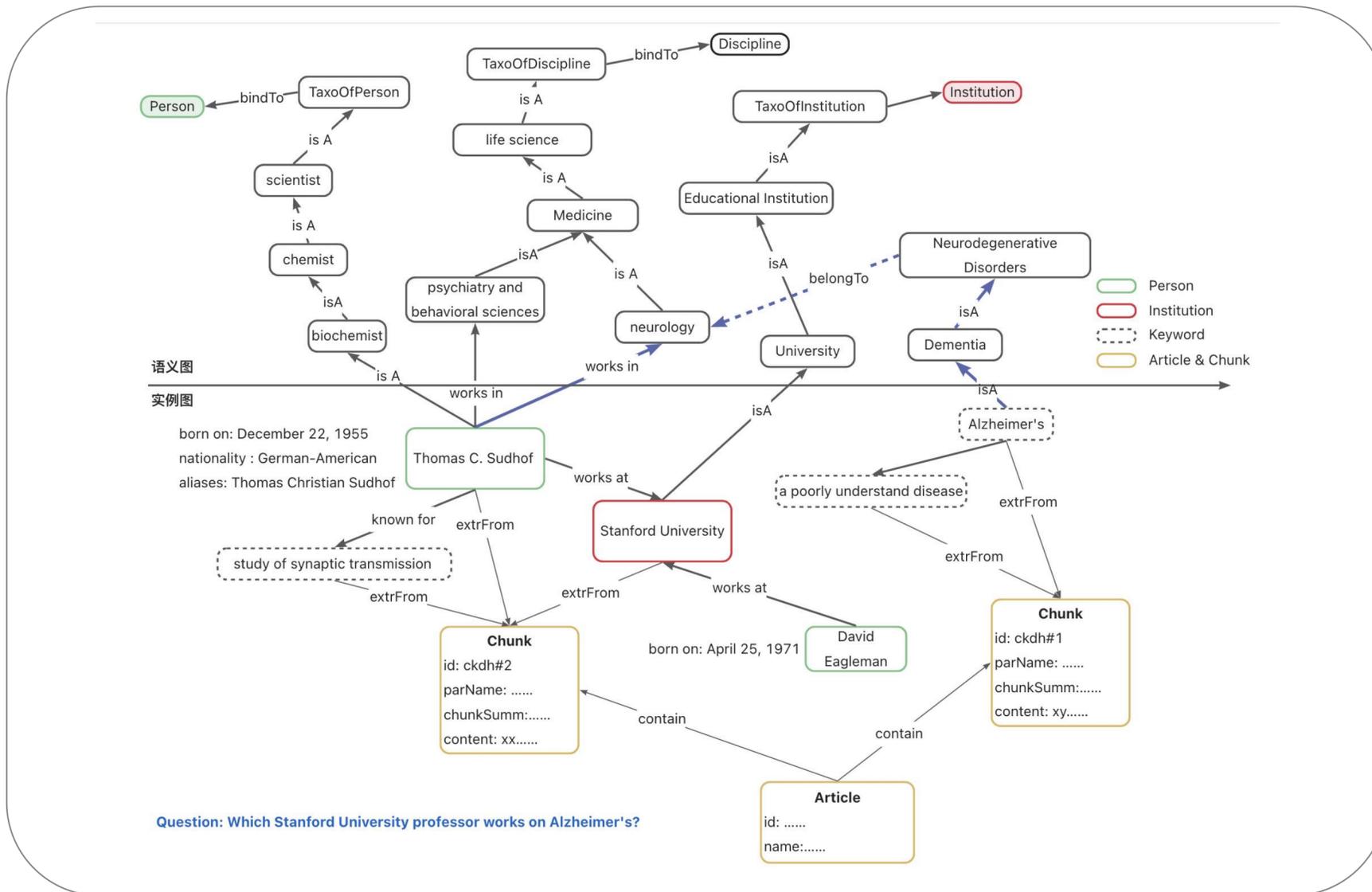
# KBQA + HippoRAG => KAG

## KGs + Docs => R&R + G



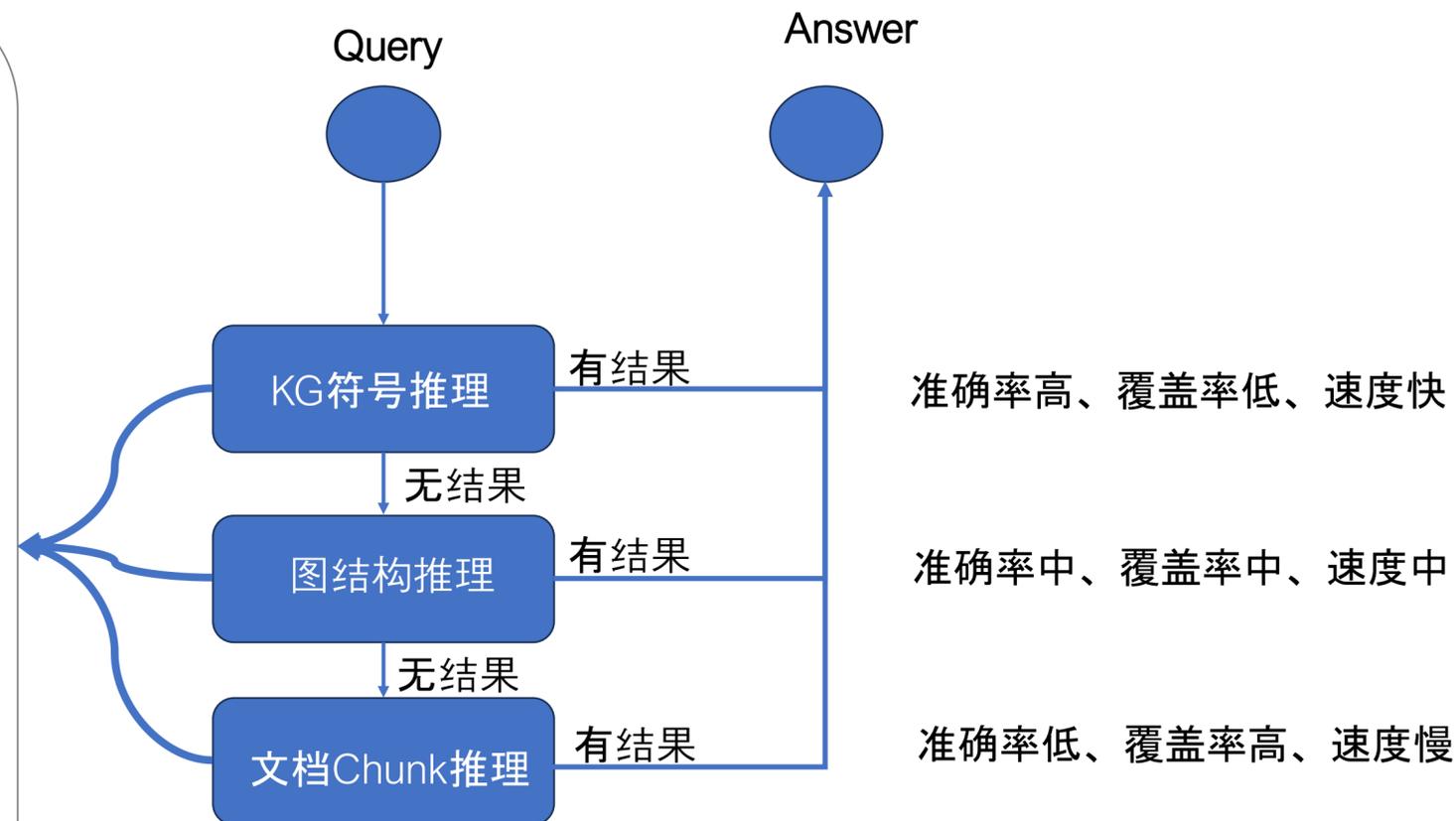
KBQA与HippoRAG的启发和借鉴

以推理(Reasoning&Retrieval)为始、LLM生成(Generation)为终



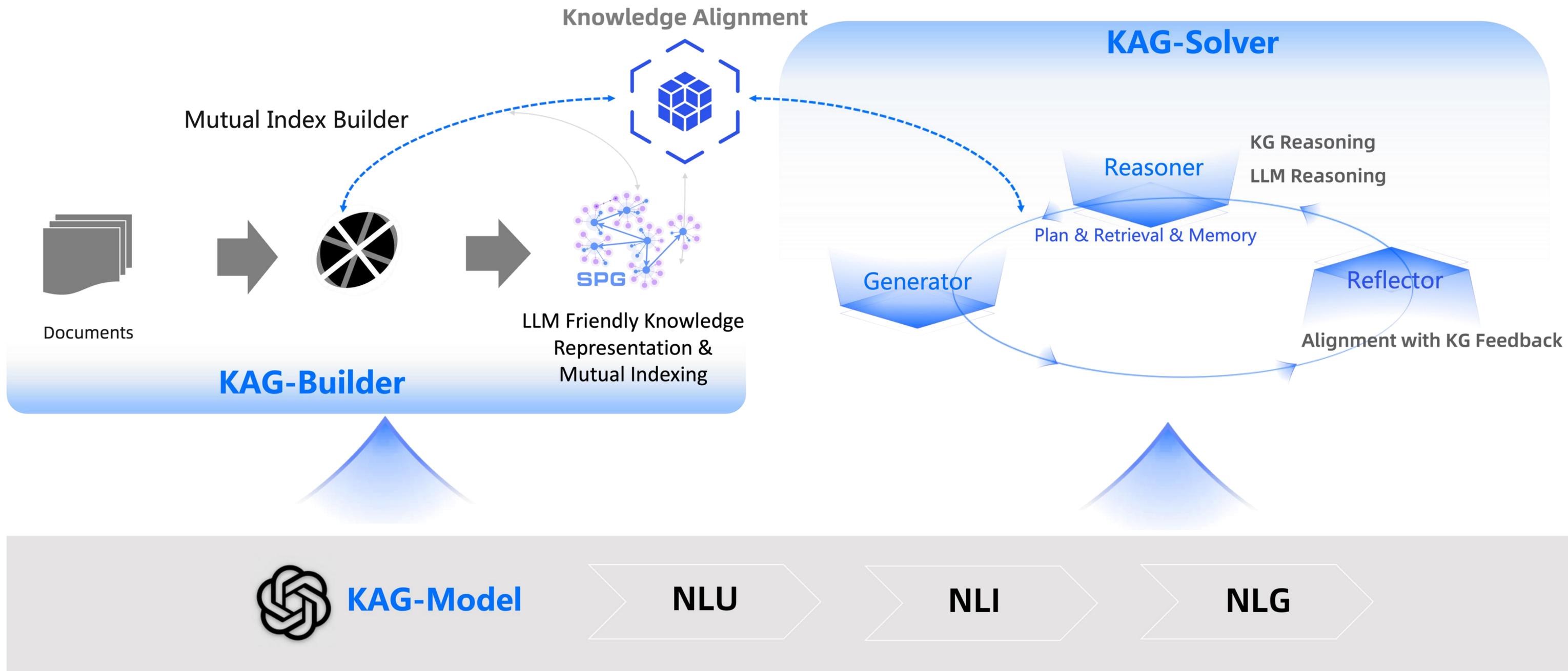
结构化知识构建

开放信息抽取

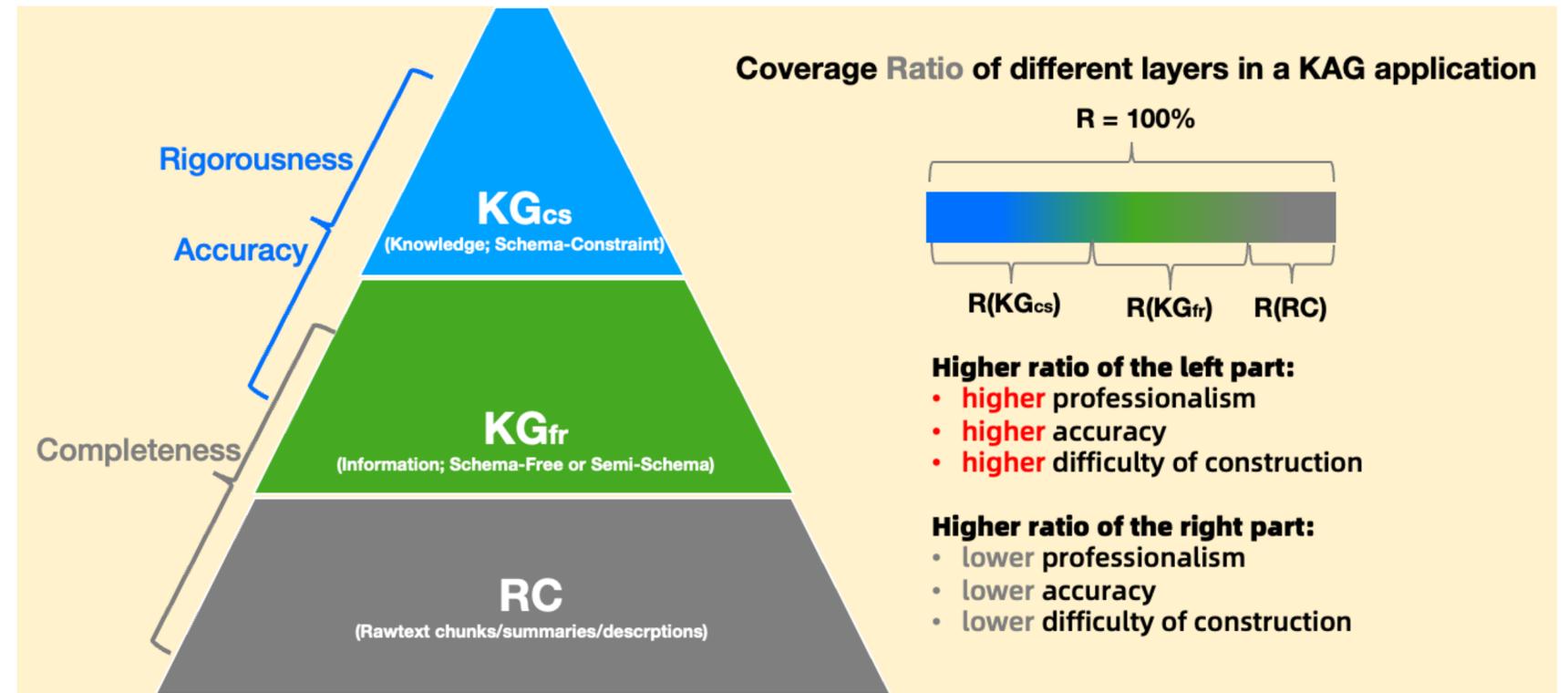
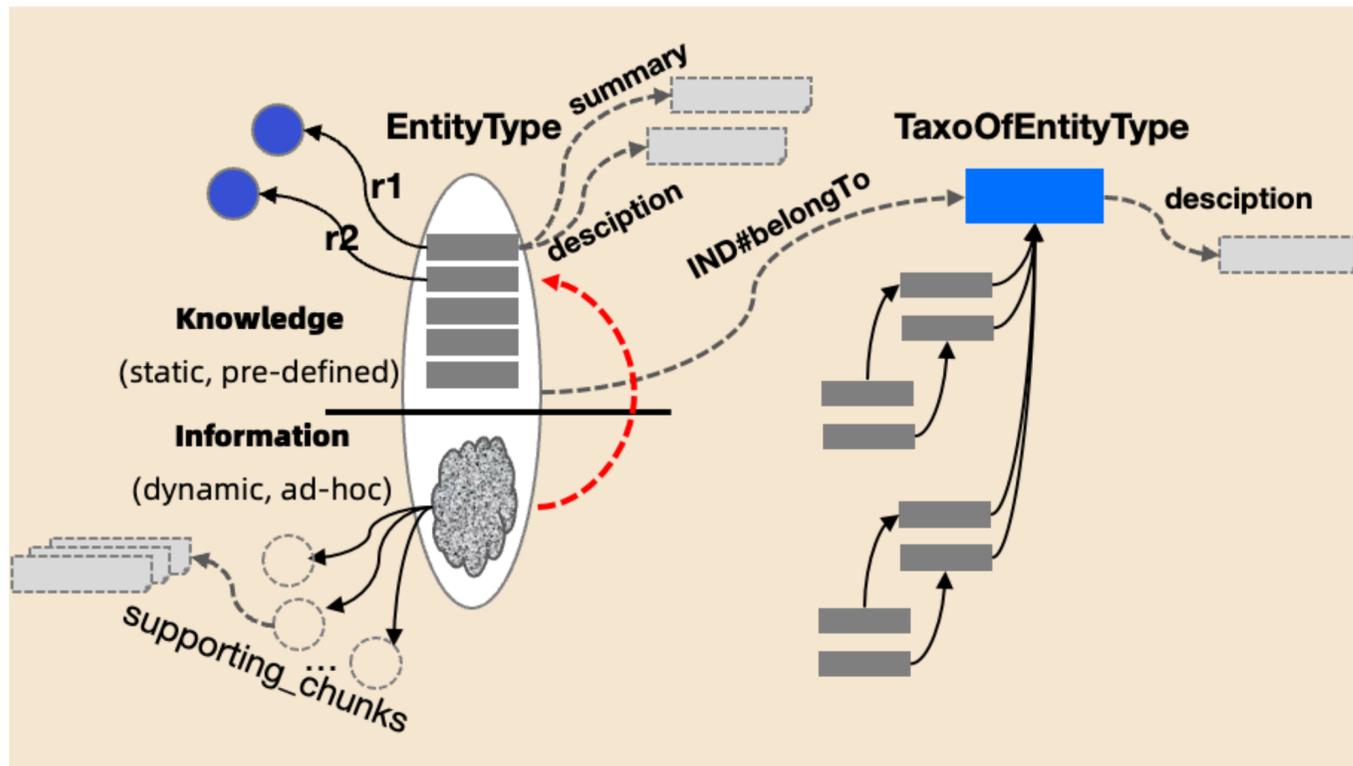


- 通过知识问答提升逻辑推理能力
- 通过信息抽取来弥补知识稀疏性
- 通过开放抽取来降低落地的门槛

逻辑符号引导的分层推理与检索



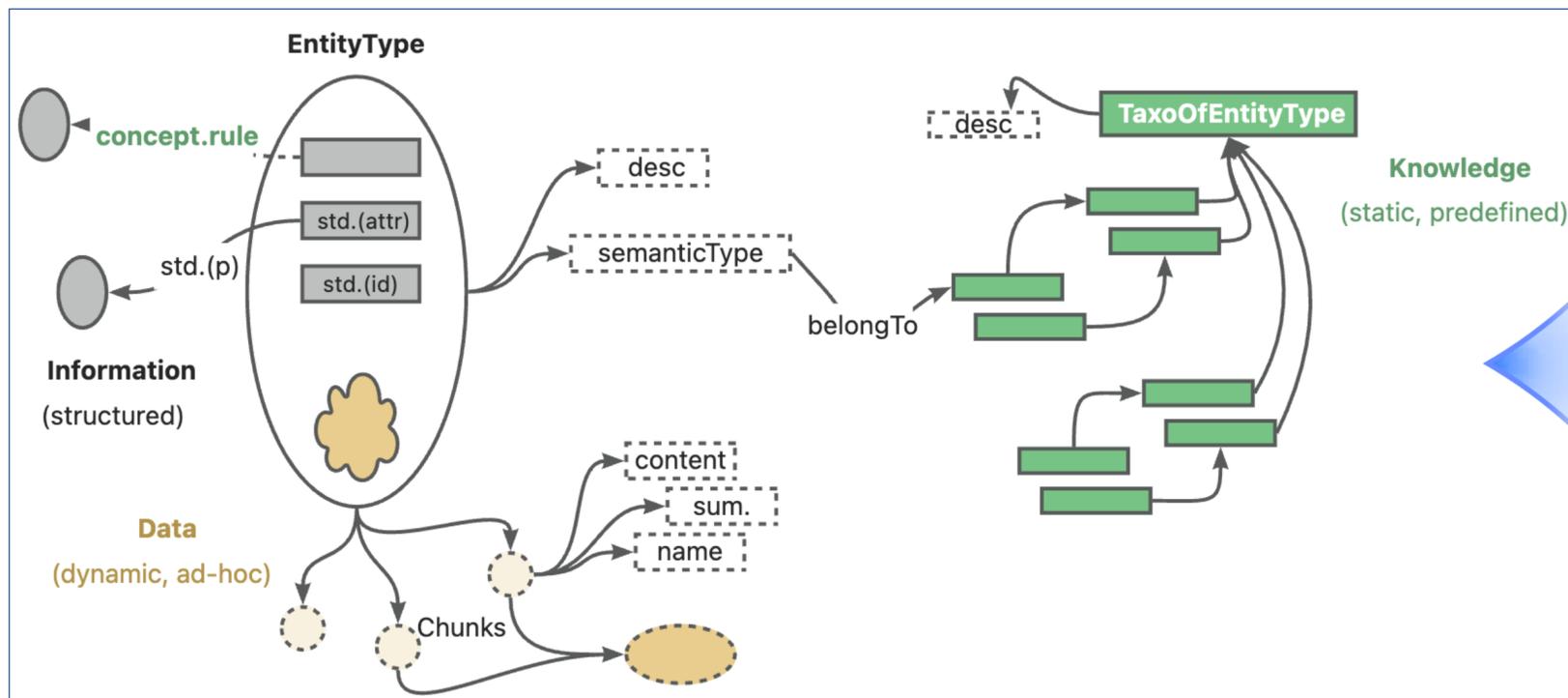
## 关键升级 1



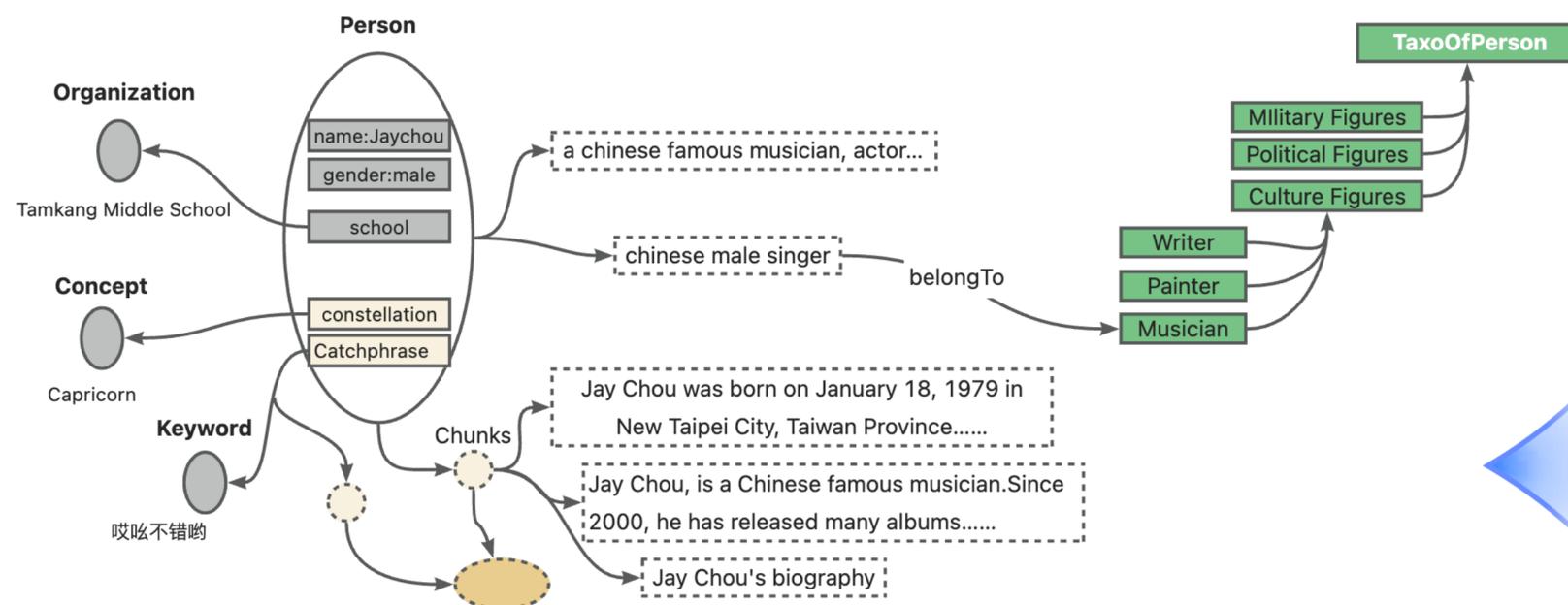
- 升级SPG为面向大模型友好的知识表示LLMFriSPG
- 兼容强Schema**专业知识**和弱Schema**开放信息**
- 图结构知识与文本知识的**互索引结构**
- 专业领域可平滑调节的**专业决策与信息检索，丰富知识完备性**

## 关键升级 1

### KAG – Indexing Structure



### KAG – Indexing instance of Jay Chou



### default.schema

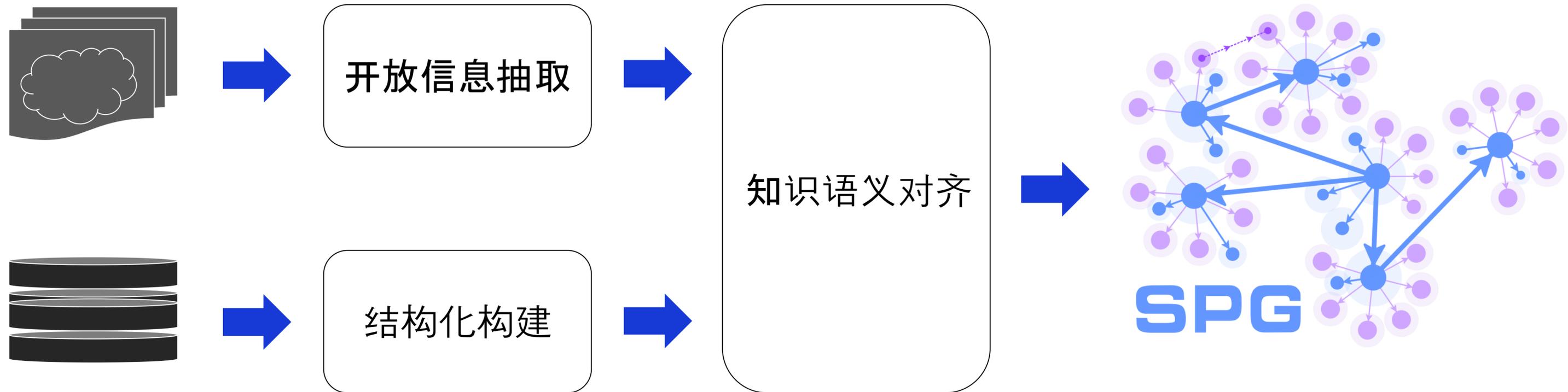
Organization(组织机构): EntityType  
 properties:  
 id(主键): Text  
 index: TextAndVector  
 name(机构名): Text  
 index: TextAndVector  
 desc(描述): Text  
 index: TextAndVector  
 semanticType(语义类型): Text

Person(人物): EntityType  
 properties:  
 id(主键): Text  
 index: TextAndVector  
 name(姓名): Text  
 index: TextAndVector  
 desc(描述): Text  
 index: TextAndVector  
 school(毕业院校): Organization  
 gender(性别): Text  
 semanticType(语义类型): Text

Works(作品): EntityType  
 Concept(概念): EntityType  
 GeoLocation(地理位置): EntityType

.....  
 Chunks(文章段落): EntityType  
**Others(其它): EntityType**

## 关键升级 1



大幅降低知识图谱的构建门槛

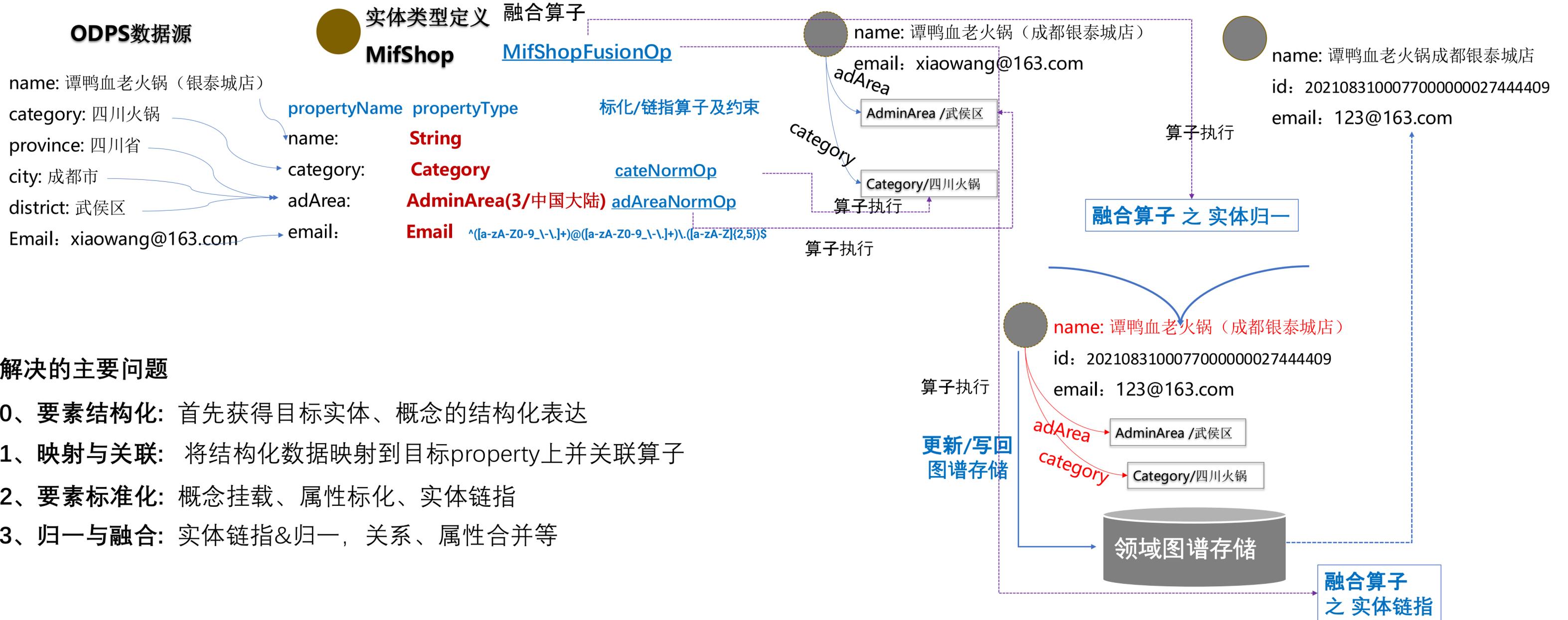
基于语义对齐平衡**信息抽取**(低门槛)与**专业构建**  
通过语义对齐缓解开放信息抽取引入的噪声问题

## 关键升级 1

Step1: mapping

Step2: 属性标化/实体链指

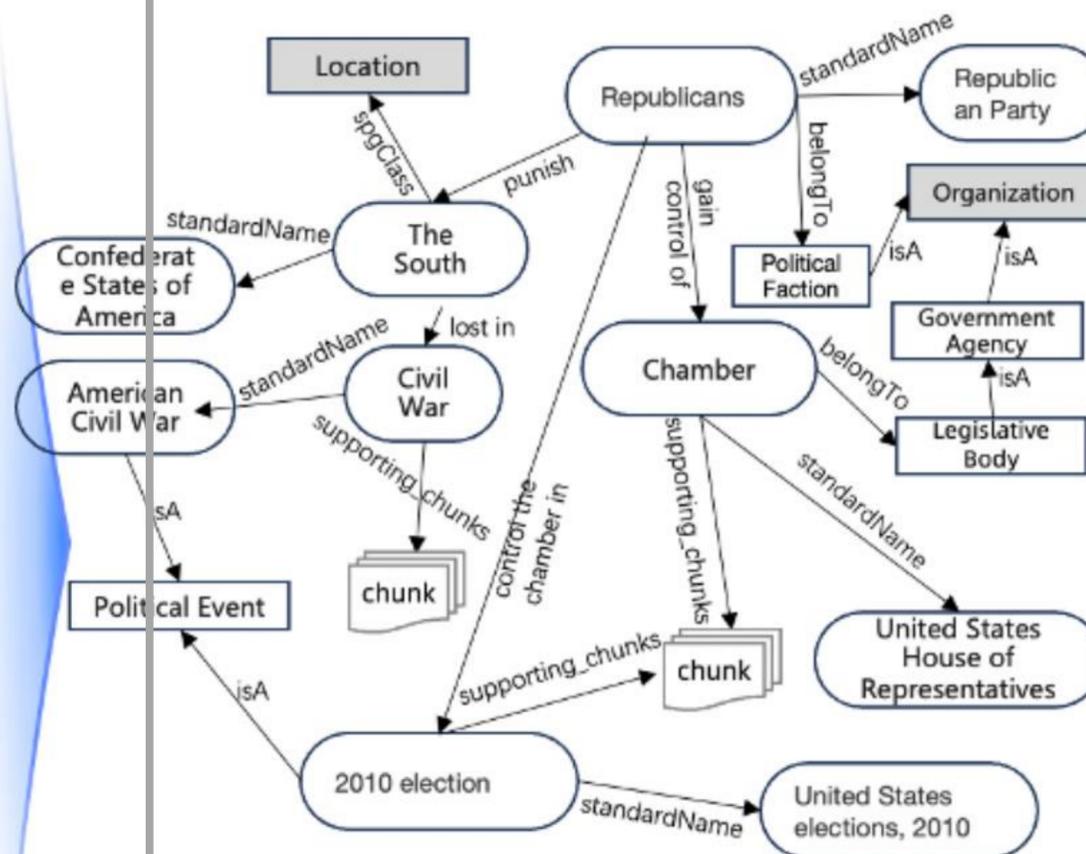
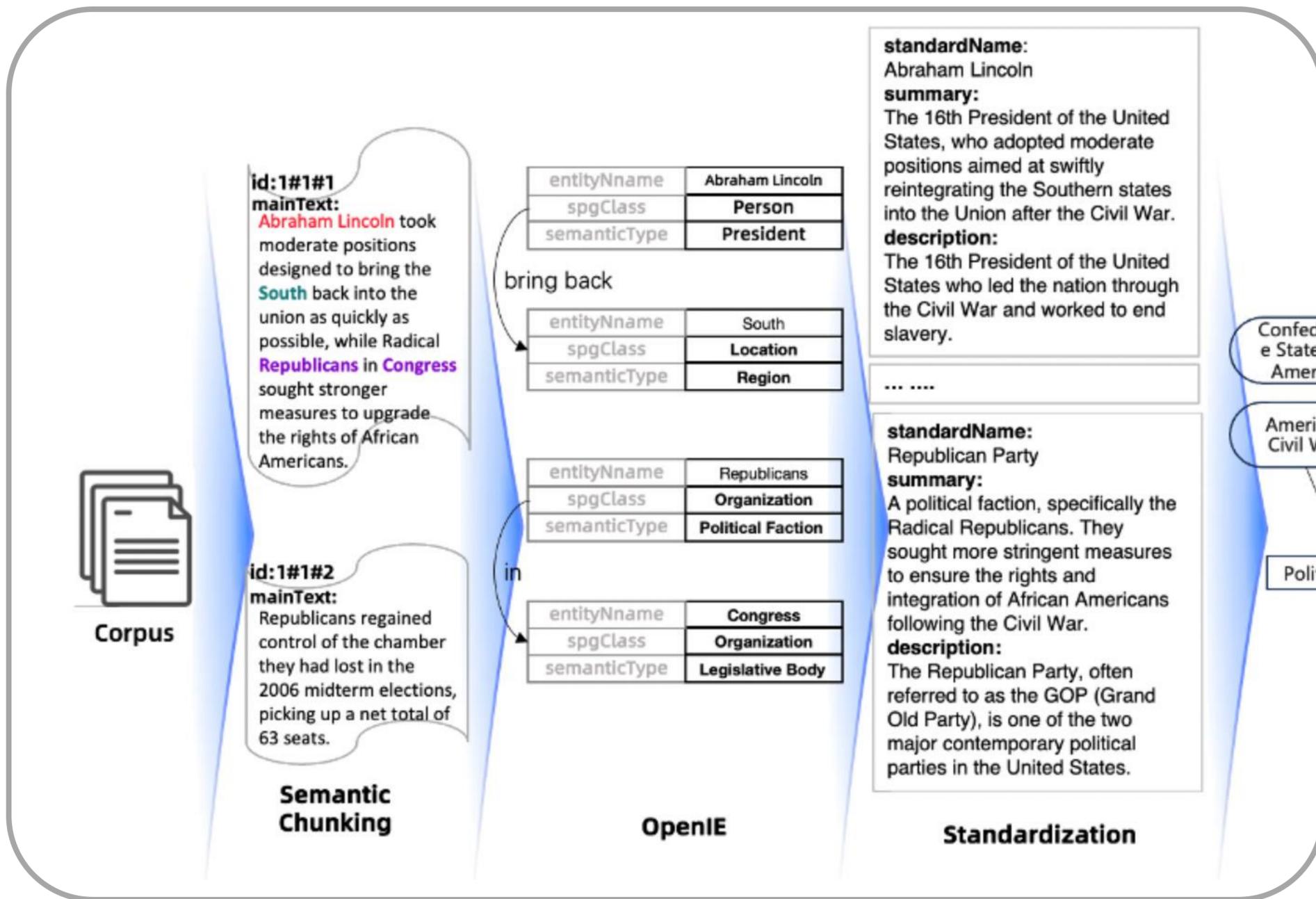
Step3: 实体融合



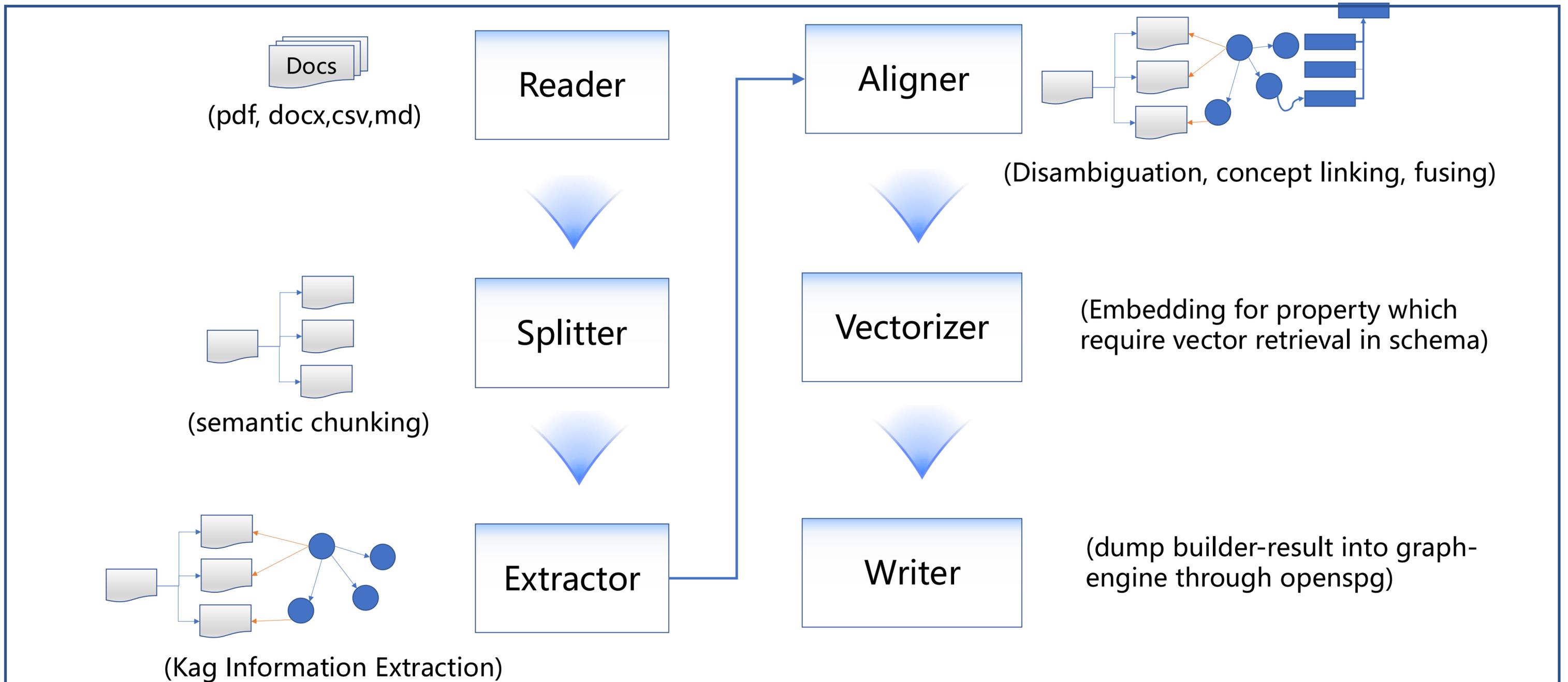
### 解决的主要问题

- 0、要素结构化: 首先获得目标实体、概念的结构化表达
- 1、映射与关联: 将结构化数据映射到目标property上并关联算子
- 2、要素标准化: 概念挂载、属性标化、实体链指
- 3、归一与融合: 实体链指&归一, 关系、属性合并等

## 关键升级 1



## 关键升级 1



## Q&A

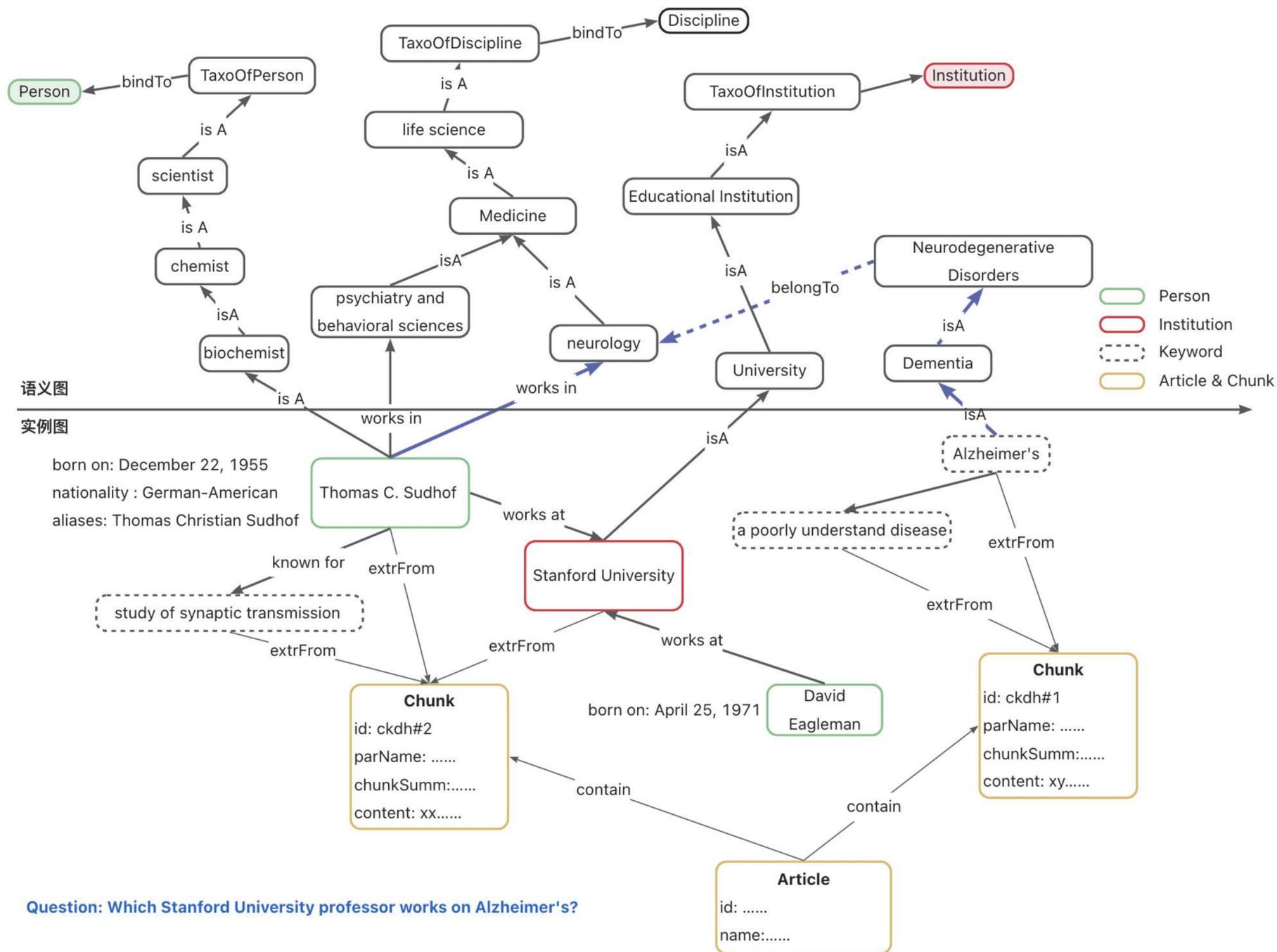
- KAG为什么要同时支持OpenIE 和 结构化构建？
- KAG为什么定义统一的知识表示和存储模型？

## 关键升级 2

概念

实例

Chunk



## 关键升级 2



Corpus

**id:1#1#1**  
**mainText:**  
Abraham Lincoln took moderate positions designed to bring the South back into the union as quickly as possible, while Radical Republicans in Congress sought stronger measures to upgrade the rights of African Americans.

**id:1#1#2**  
**mainText:**  
Republicans regained control of the chamber they had lost in the 2006 midterm elections, picking up a net total of 63 seats.

Semantic Chunking

entityNname	Abraham Lincoln
spgClass	Person
semanticType	President

bring back

entityNname	South
spgClass	Location
semanticType	Region

in

entityNname	Republicans
spgClass	Organization
semanticType	Political Faction

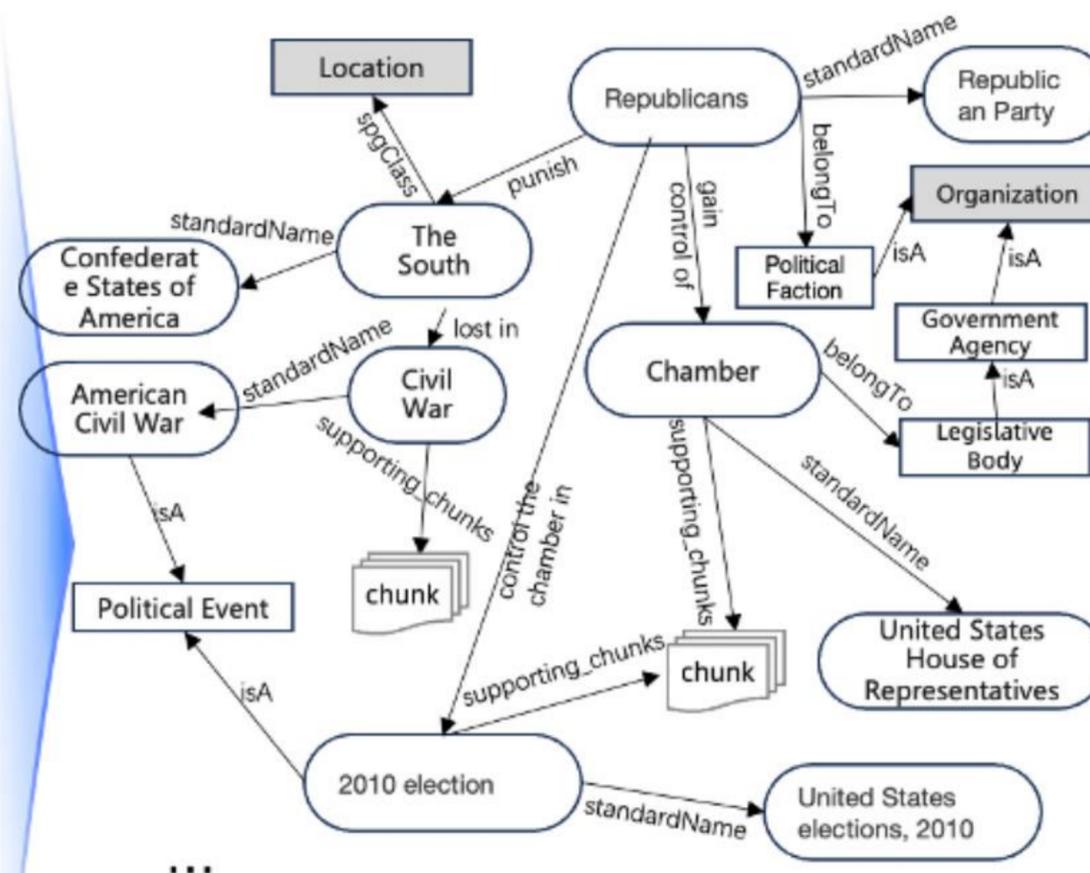
entityNname	Congress
spgClass	Organization
semanticType	Legislative Body

OpenIE

**standardName:**  
Abraham Lincoln  
**summary:**  
The 16th President of the United States, who adopted moderate positions aimed at swiftly reintegrating the Southern states into the Union after the Civil War.  
**description:**  
The 16th President of the United States who led the nation through the Civil War and worked to end slavery.

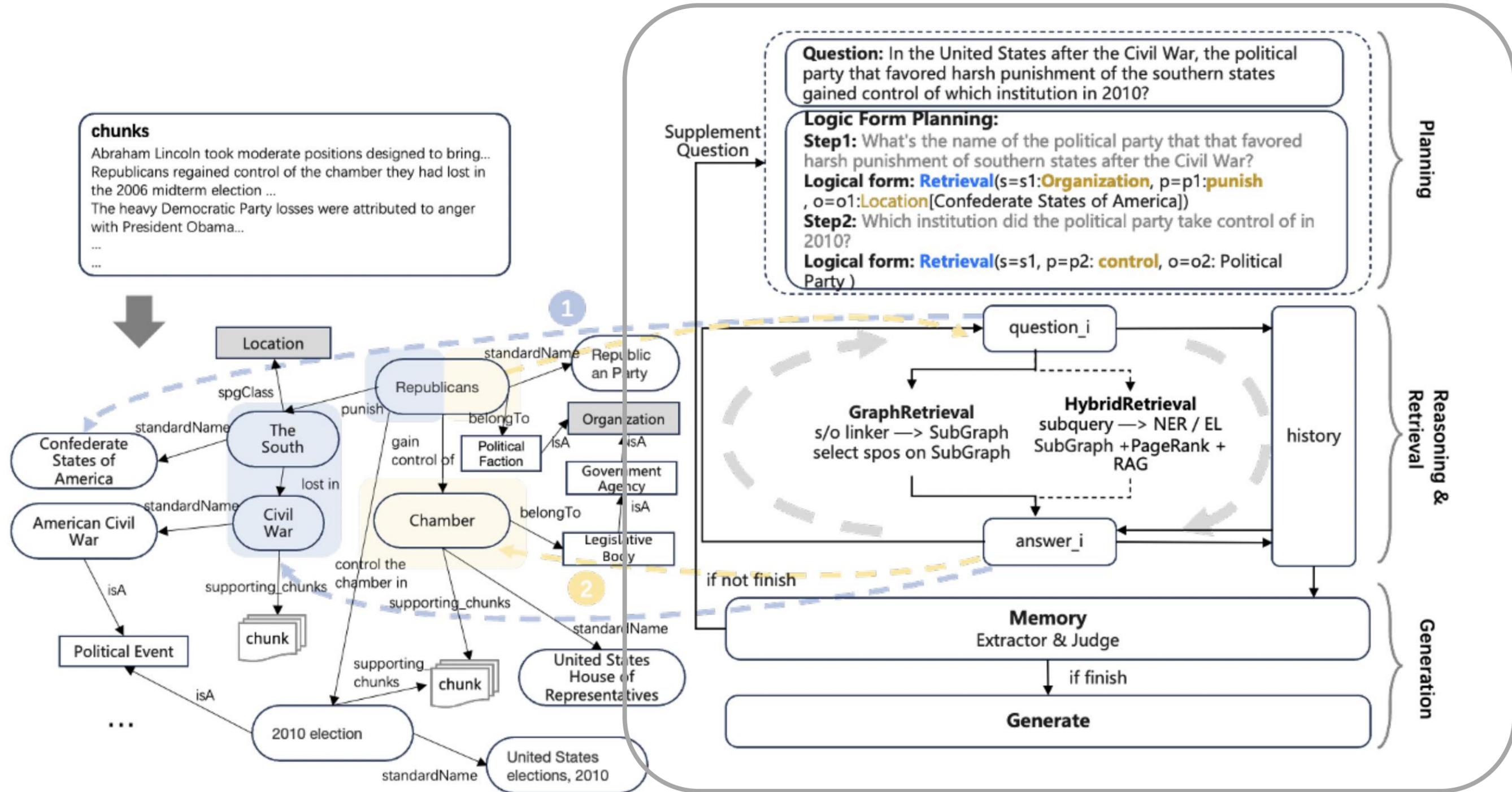
**standardName:**  
Republican Party  
**summary:**  
A political faction, specifically the Radical Republicans. They sought more stringent measures to ensure the rights and integration of African Americans following the Civil War.  
**description:**  
The Republican Party, often referred to as the GOP (Grand Old Party), is one of the two major contemporary political parties in the United States.

Standardization



Concept Completion and KG Construction

## 关键升级 2



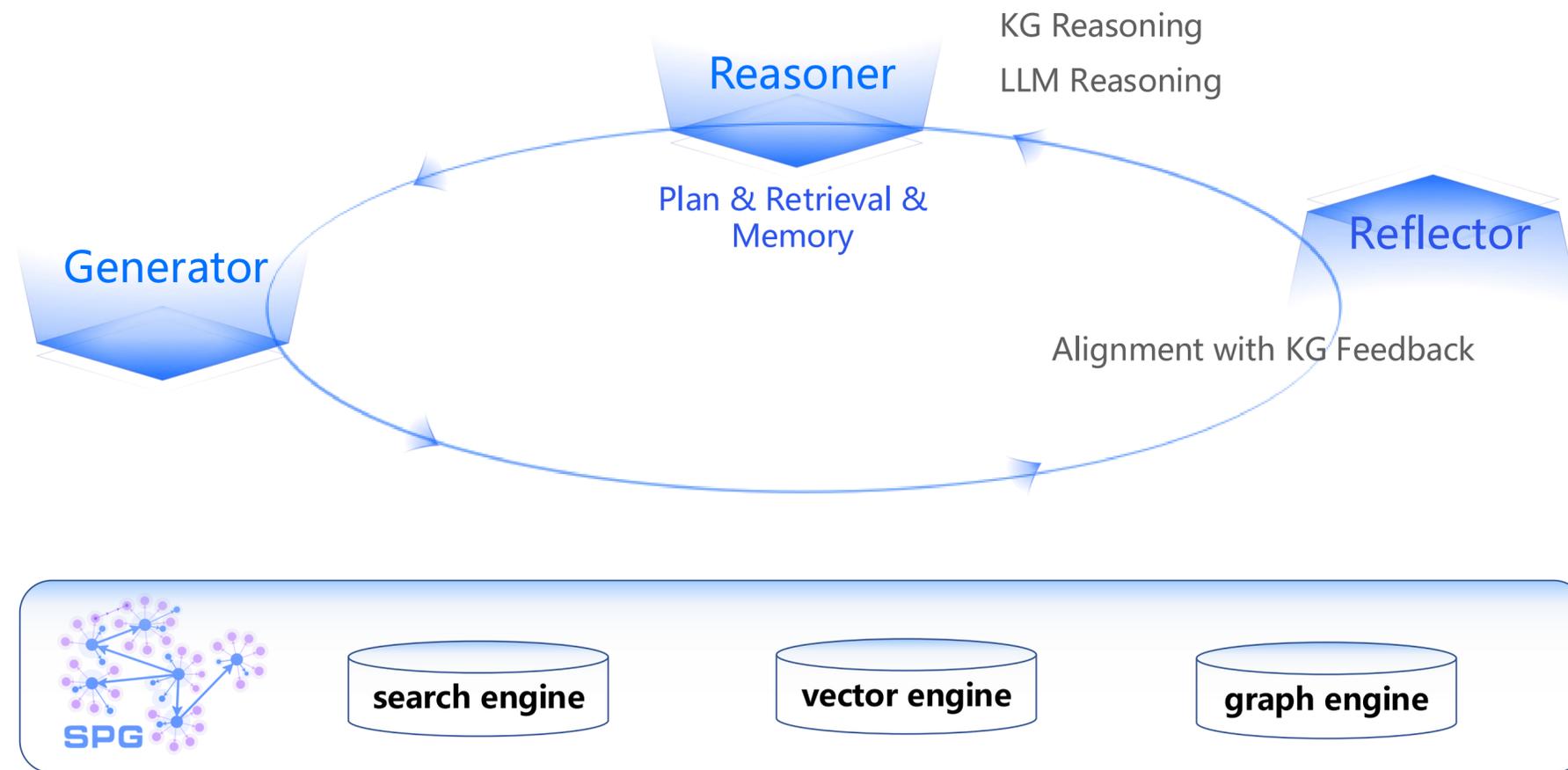
## 关键升级 2

Formal Expression	Description	Example
$\langle var1, synonym, var2 \rangle$	A <i>synonym</i> relation means that a word or phrase <i>var2</i> that has the same or nearly the same meaning as another word or phrase <i>var1</i> in the same language and given context.	<i>Fast</i> is a synonym of <i>quick</i> .
$\langle var1, isA, var2 \rangle$	An <i>isA</i> relation means that a hypernym <i>var2</i> that is more generic or abstract than a given word or phrase <i>var1</i> and encompasses a broader category that the given word belongs to.	<i>Car</i> isA <i>Vehicle</i> .
$\langle var1, isPartOf, var2 \rangle$	An <i>isPartOf</i> relation means that something <i>var1</i> is a component or constituent of something <i>var2</i> larger. This relation shows that an item is a part of a bigger whole.	<i>Wheel</i> isPartOf <i>car</i> .
$\langle var1, contains, var2 \rangle$	A <i>contains</i> relation means that something <i>var1</i> includes or holds <i>var2</i> , something else within it. This indicates that one item has the other as a subset or component.	<i>Library</i> contains <i>books</i> .
$\langle var1, belongTo, var2 \rangle$	An <i>belongTo</i> relation means that something <i>var1</i> is an instance of concept <i>var2</i> .	<i>Chamber</i> belongTo <i>Legislative Body</i> .
$\langle var1, causes, var2 \rangle$	A <i>causes</i> relation means that one event or action <i>var1</i> brings about another <i>var2</i> . This indicates a causal relation where one thing directly results in the occurrence of another.	<i>Fire</i> causes <i>smoke</i> .

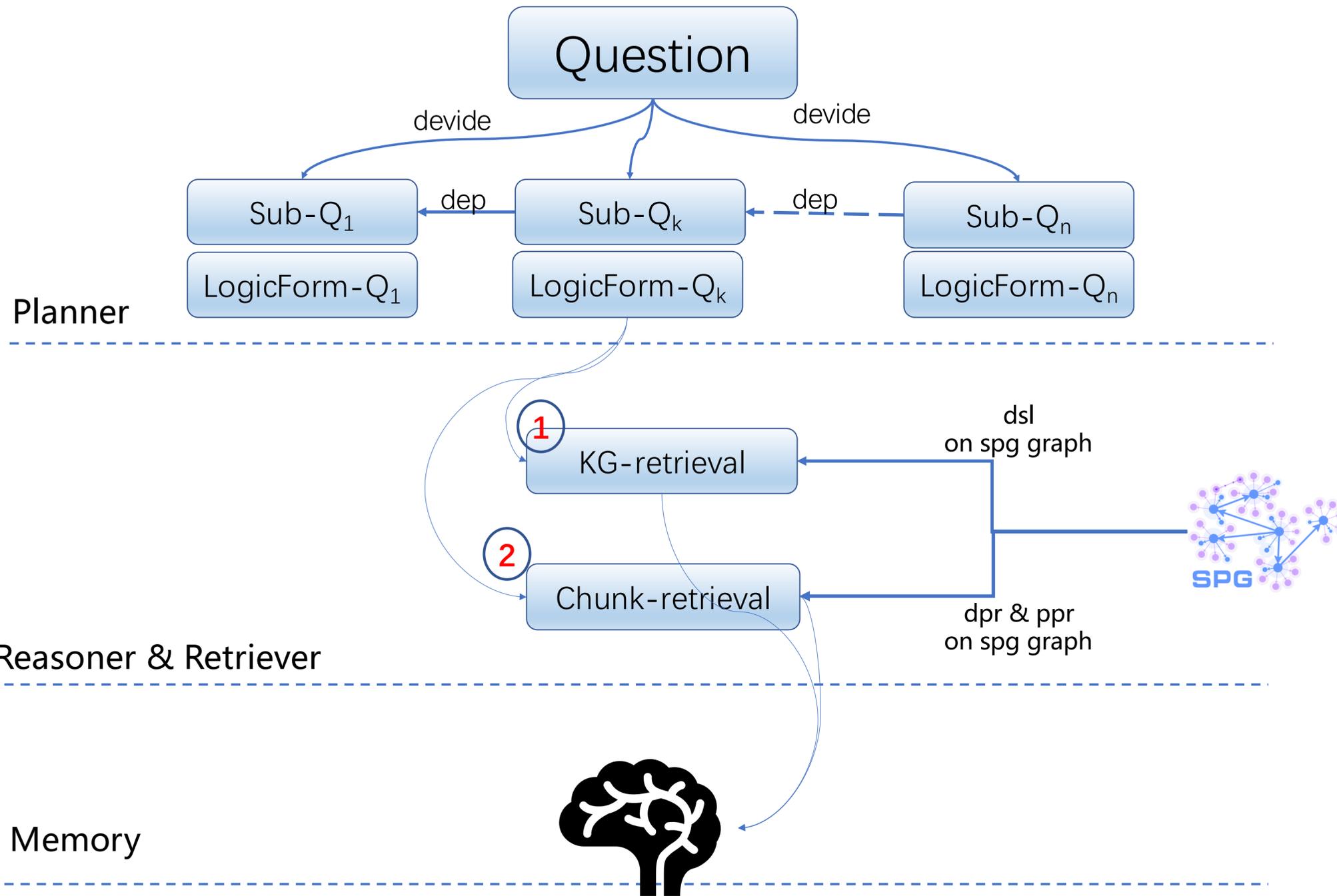
## Q&A

- 语义对齐到底解决了什么问题？
- 怎么实现领域知识的对齐？

## 关键升级 3



## 关键升级 3



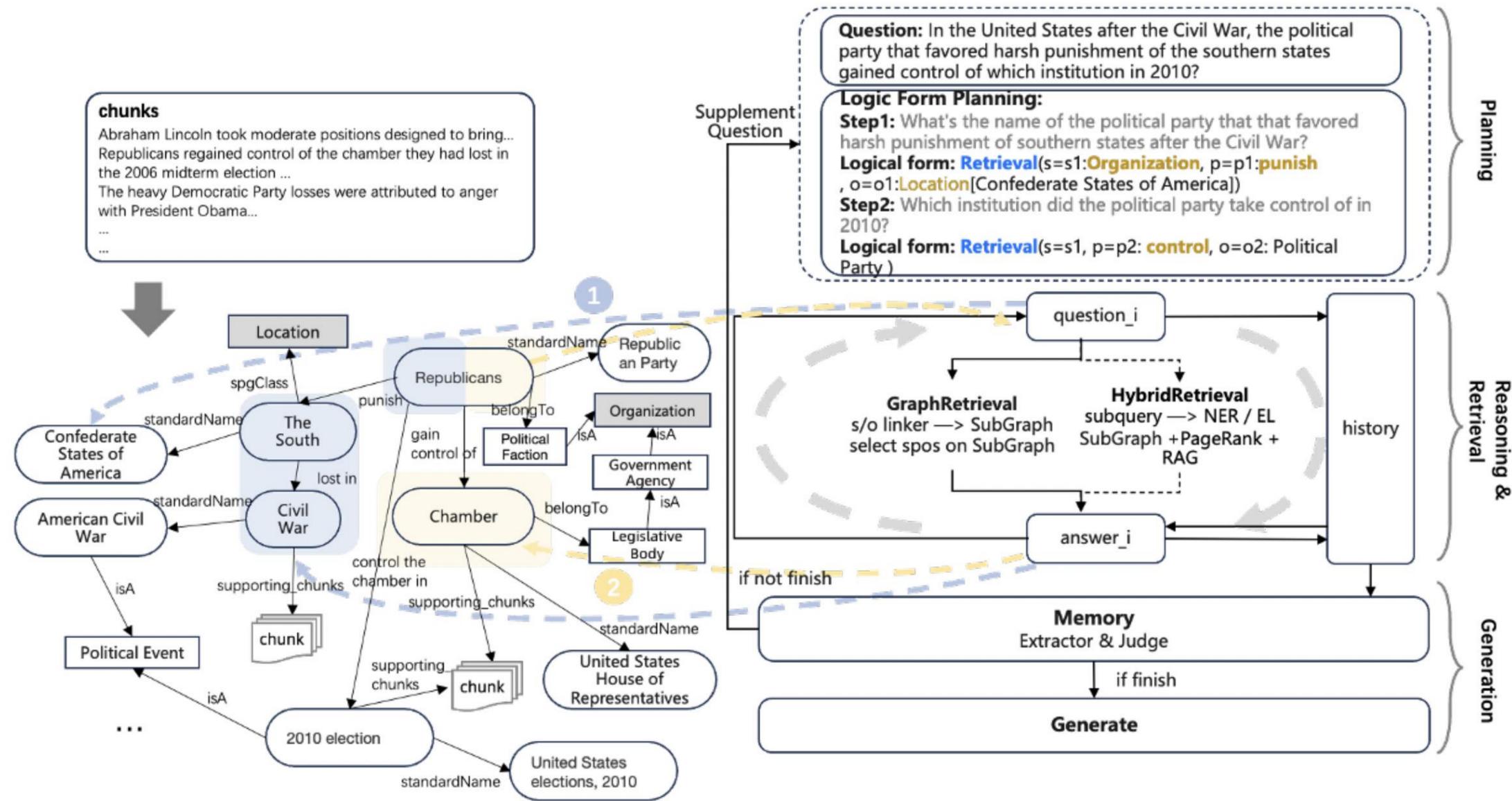
## 关键升级 3

### Algorithm 1 Logical Form Solver

```

1:  $memory \leftarrow []$ 
2:  $query_{cur} \leftarrow query$ 
3: for  $round \in (0, n)$  do
4:    $lf_{list} \leftarrow LFPlanner(query_{cur})$ 
5:    $history \leftarrow []$ 
6:   for  $lf \in lf_{list}$  do
7:      $lf_{subquery}, lf_{func} \leftarrow lf$ 
8:      $retrievals_{sub}, answer_{sub} \leftarrow Reasoner(lf_{subquery}, lf_{func})$ 
9:      $history.append([lf_{subquery}, retrievals_{sub}, answer_{sub}])$ 
10:  end for
11:   $memory \leftarrow Memory(query, history)$ 
12:  if not  $Judge(query, memory)$  then
13:     $query_{cur} \leftarrow SupplyQuery(query, memory)$ 
14:  end if
15: end for
16:  $answer \leftarrow Generator(query, memory)$ 
17: return  $answer$ 

```



符号推理 与 参数推理融合图谱符号推理、向量检索及大模型推理

## Q&A

- Logical form的问题拆解会有错误累加问题么？
- 从代码来看Logical form不是语义完备的， Logical form后面怎么发展？
- Logical form和KGDSL之间的边界和转换逻辑是什么？

# 基于KAG的业务应用

## 知识问答

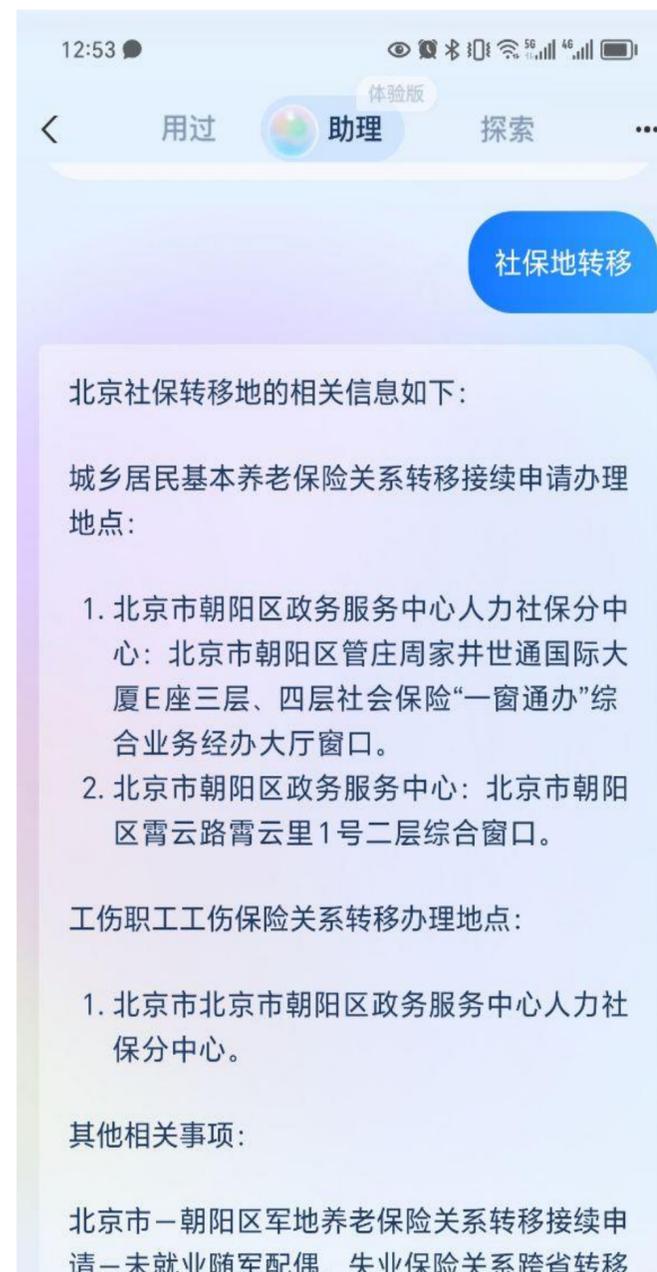
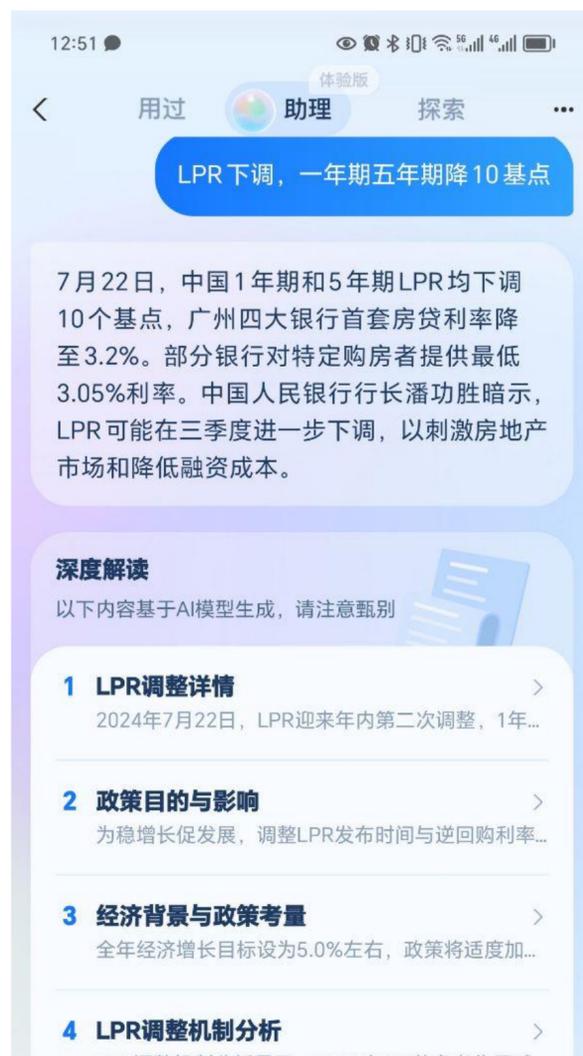
### 政务办事问答

### 医疗健康问答

## 分析写作

### 热点事件解读

### 银行风险分析



保定市提取公积金需要哪些材料

购买自住住房提取住房公积金

不全，提取公积金的12种不同情况

生育险在哪里看

生育津贴支付

检索错误，参保人员参保信息查询

	有果 准确率	召回率
Basic RAG	0.55	0.37
KAG(内部)	0.91	0.71

舟山市怎么查房产证

未找到相关信息

遗漏，不动产权属证明网上查询

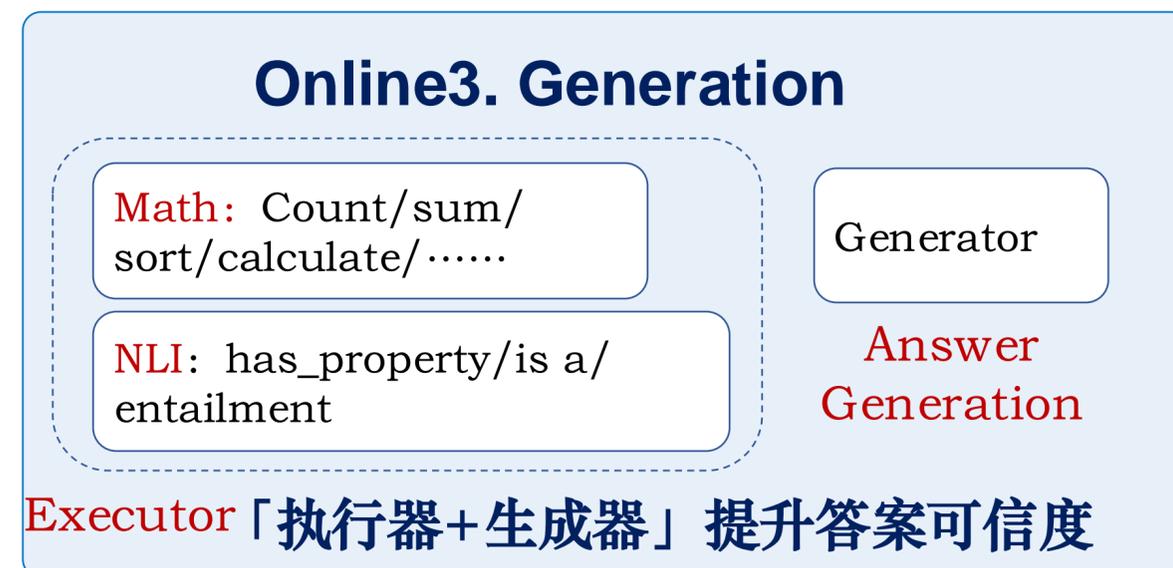
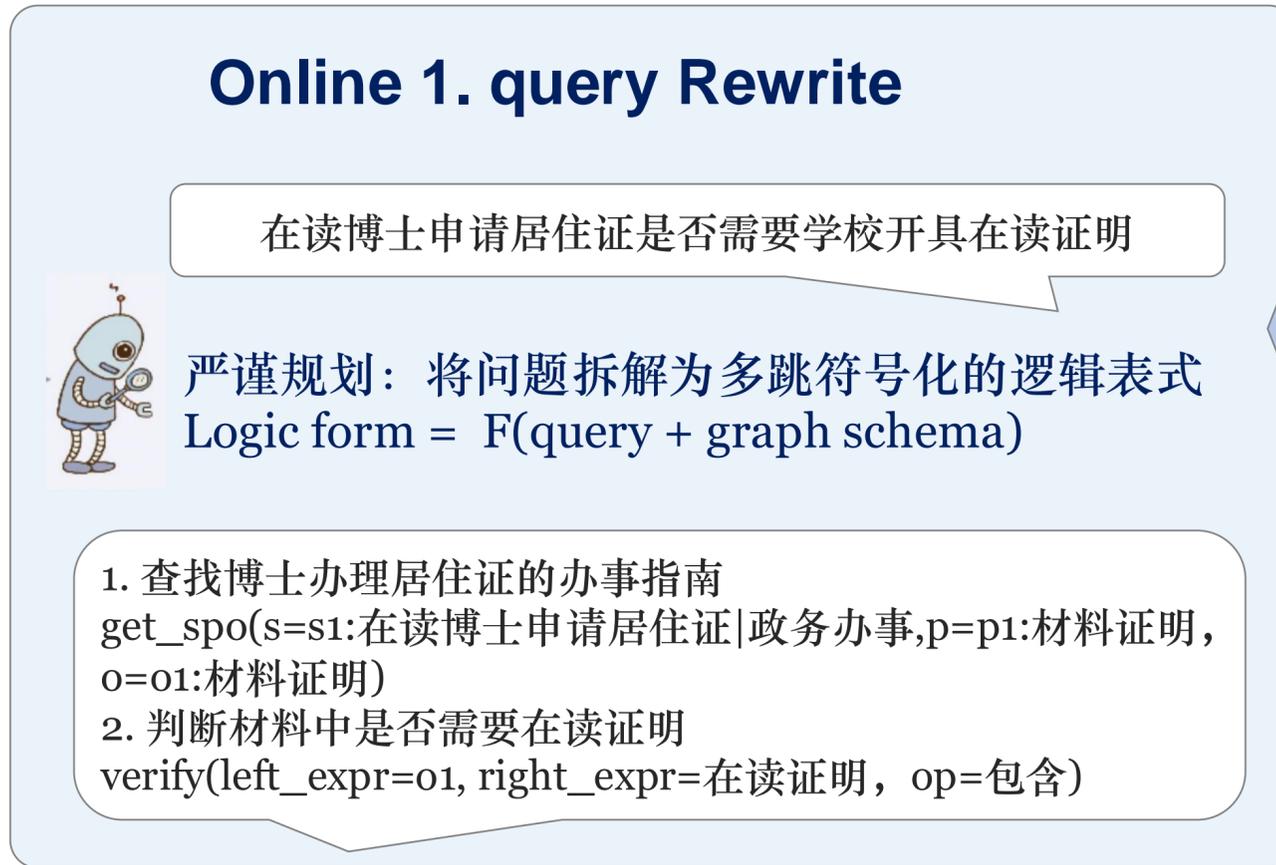
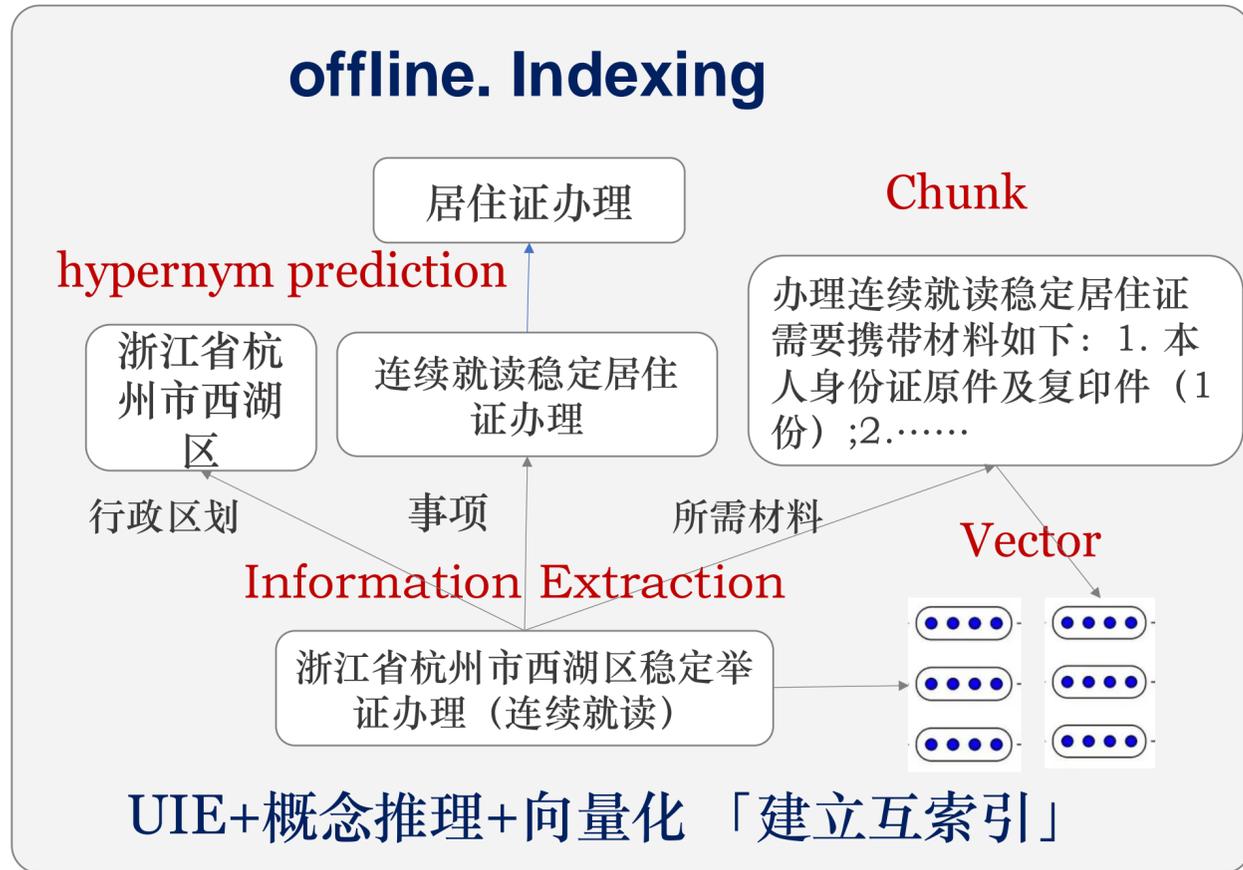
社保月缴费多少

职工参保登记

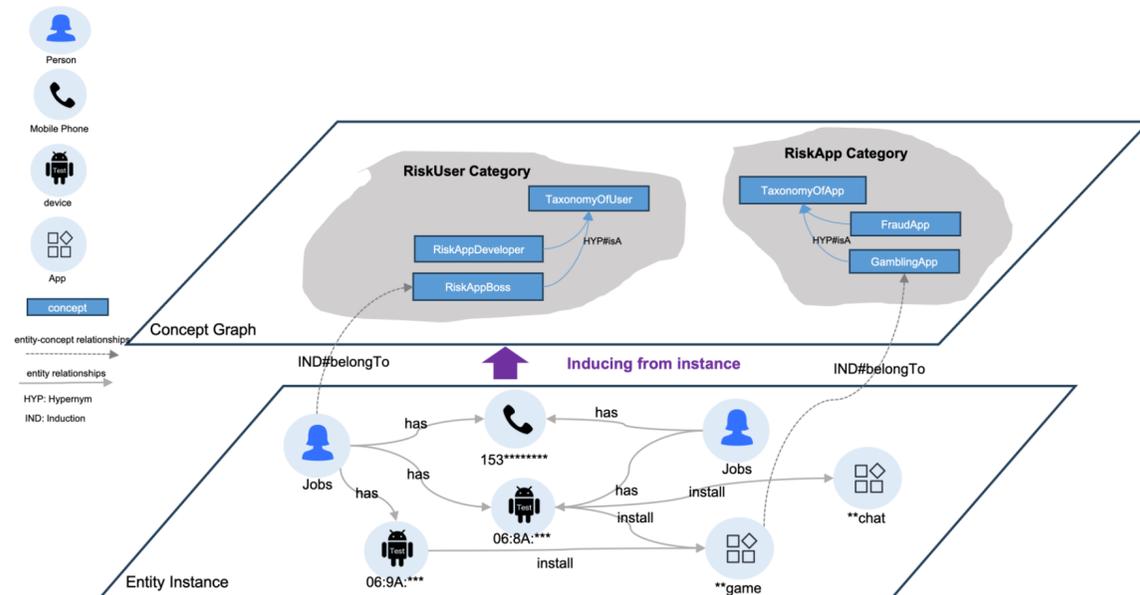
检索错误，没有事项

我买房申请了500万公积金贷款，计划20年还清，当前月利率是0.02，每个月应该还多少钱

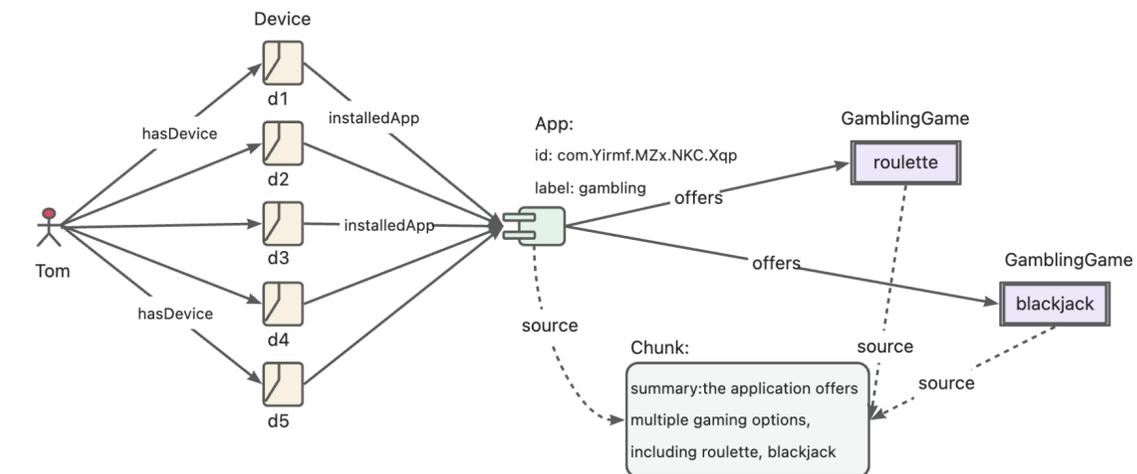
已知公积金贷款月供计算公式为[贷款本金×月利率×(1+月利率)<sup>还款月数</sup>]÷[(1+月利率)<sup>还款月数</sup>-1]  
算式：500万×0.02×(1+0.02)<sup>(20×12)</sup>÷[(1+0.02)<sup>(20×12)</sup>-1]



黑产图谱，通过提升黑产特征挖掘、完善黑产团伙刻画、推送线索助力司法部门线下打击等手段，降低体系内赌博电诈风险。



## 典型风险模式



## 1 定义"赌博App" 规则

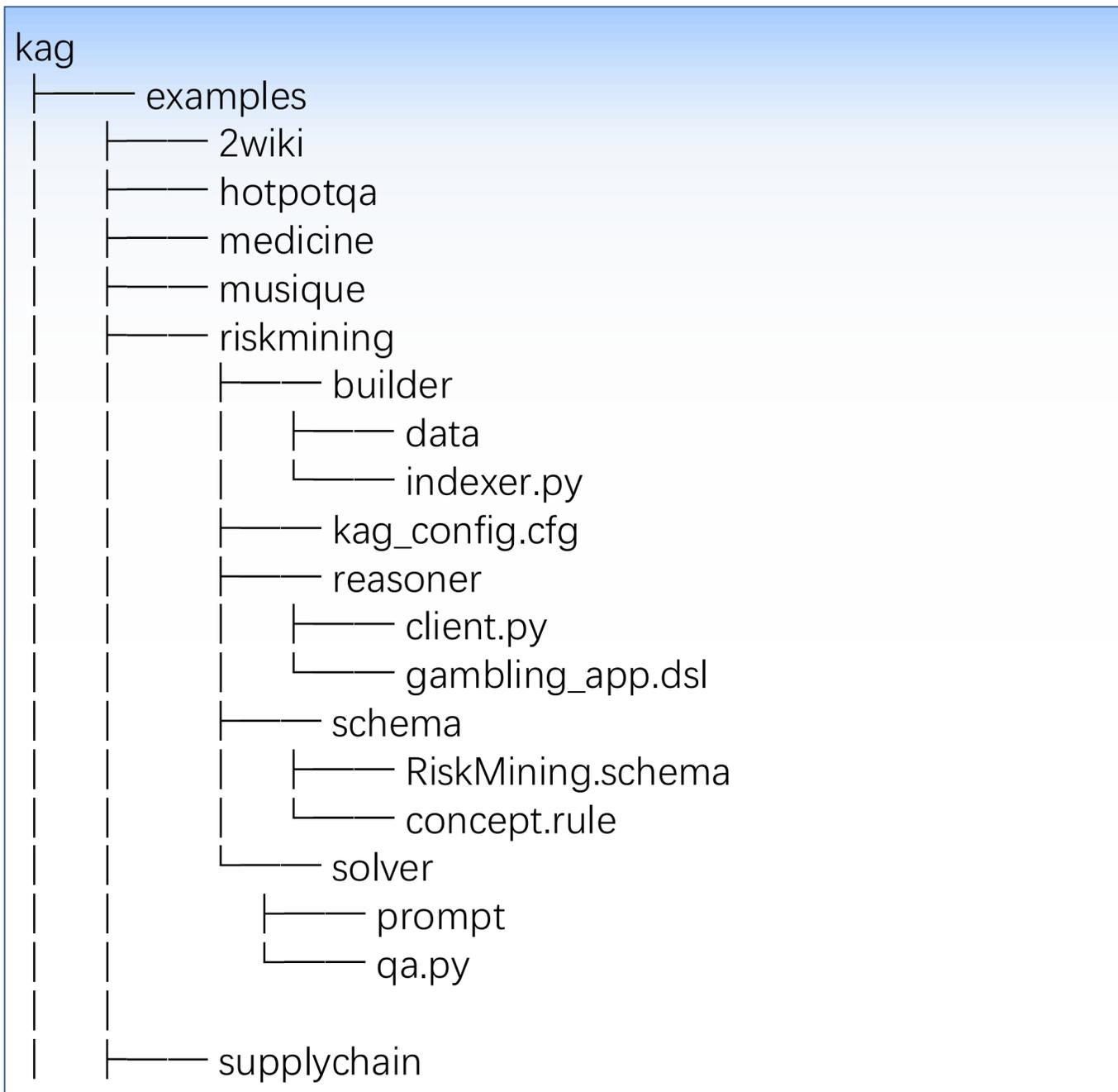
```
define riskAppTaxo rule
  Define (s:App)-[p:belongTo]->(o:`TaxOfRiskApp`/`GamblingApp`) {
    Structure {
      (s)
    }
    Constraint {
      R1("risk label marked as gambling") s.riskMark like "%Gambling%"
    }
  }
```

## 2 定义"App开发者" 规则

```
define app developer rule
  Define (s:Person)-[p:developed]->(o:App) {
    Structure {
      (s)-[:hasDevice]->(d:Device)-[:install]->(o)
    }
    Constraint {
      deviceNum = group(s,o).count(d)
      R1("device installed same app"): deviceNum > 5
    }
  }
```

The screenshot shows the OpenSPG/KAG user interface. At the top, there are navigation tabs for '首页', 'RiskMining', '知识库管理', and '知识库问答'. A '新建查询对话' button is prominent. Below it, a '历史会话' section shows a previous query: '裘\*\*是否有风险'. The main area displays a Q&A session with the question '裘\*\*是否有风险' and two sub-questions: '子问题1: 查询裘\*\*的分类' and '子问题2'. The '问题回答' section shows the question and the answer: '答案: 赌博App开发者'. On the right, a 'SPO Retriever' panel shows the logic form expression 'get\_spo(s=s1:自然人[裘\*\*], p:...' and the retrieved result: '['(裘\*\* belongTo 赌博App开发者)']'.

Logical Form 2 KGDSL



Framework	Model	HotpotQA		2WikiMultiHopQA		MuSiQue	
		EM	F1	EM	F1	EM	F1
NativeRAG [35, 34]	ChatGPT-3.5	43.4	57.7	33.4	43.3	15.5	26.4
HippoRAG [12, 34]	ChatGPT-3.5	41.8	55.0	46.6	59.2	19.2	29.8
IRCoT+NativeRAG	ChatGPT-3.5	45.5	58.4	35.4	45.1	19.1	30.5
IRCoT+HippoRAG	ChatGPT-3.5	45.7	59.2	47.7	62.7	21.9	33.3
IRCoT+HippoRAG	DeepSeek-V2	51.0	63.7	48.0	57.1	26.2	36.5
<b>KAG w/ LFS<sub>ref3</sub></b>	DeepSeek-V2	<u>59.8</u>	<u>74.0</u>	<u>66.3</u>	<u>76.1</u>	<u>35.4</u>	<u>48.2</u>
<b>KAG w/ LFSH<sub>ref3</sub></b>	DeepSeek-V2	<b>62.5</b>	<b>76.2</b>	<b>67.8</b>	<b>76.2</b>	<b>36.7</b>	<b>48.7</b>

Table 8: The end-to-end generation performance of different RAG models on three multi-hop Q&A datasets. The values in **bold** and underline are the best and second best indicators respectively.

	Retriever	HotpotQA		2Wiki		MuSiQue	
		Recall@2	Recall@5	Recall@2	Recall@5	Recall@2	Recall@5
Single-step	BM25 [36]	55.4	72.2	51.8	61.9	32.3	41.2
	Contriever [37]	57.2	75.5	46.6	57.5	34.8	46.6
	GTR [38]	59.4	73.3	60.2	67.9	37.4	49.1
	RAPTOR [39]	58.1	71.2	46.3	53.8	35.7	45.3
	Proposition [40]	58.7	71.1	56.4	63.1	37.6	49.3
	NativeRAG [35, 34]	64.7	79.3	59.2	68.2	37.9	49.2
	HippoRAG [12, 34]	60.5	77.7	70.7	89.1	40.9	51.9
	Multi-step	IRCoT + BM25	65.6	79.0	61.2	75.6	34.2
IRCoT + Contriever		65.9	81.6	51.6	63.8	39.1	52.2
IRCoT + NativeRAG		<u>67.9</u>	82.0	64.1	74.4	41.7	53.7
IRCoT + HippoRAG		67.0	<u>83.0</u>	<b>75.8</b>	<b>93.9</b>	<u>45.3</u>	<u>57.6</u>
<b>KAG</b>		<b>72.8</b>	<b>88.8</b>	<u>65.4</u>	<u>91.9</u>	<b>48.5</b>	<b>65.7</b>

Table 9: The performance of different retrieval models on three multi-hop Q&A datasets

框架	适用场景	量化指标 (hotpotqa 数据集)	特点
GraphRAG( MS)	摘要生成类任务 (评估方式: 全面性、多样性、赋权性)	EM: 0 F1: 0.053	<ul style="list-style-type: none"> <li>通过层次聚类实现段落摘要的逐级生成, 更关注答案生成的可理解性、完整性、多视角</li> <li>多跳问答等评测集量化指标较差, 未提供逻辑符号推理的能力</li> </ul>
HippoRAG	事实问答类任务 (评估方式::EM、F1)	EM: 0.457 F1 : 0.592	<ul style="list-style-type: none"> <li>通过rdf 抽取 + 语义相似拉边, 完成图谱构建</li> <li>问答阶段, 通过dpr + ppr 实现Chunk 召回</li> <li>未利用语义、逻辑、符号等图谱技术栈</li> </ul>
LightRAG	摘要生成类任务 (评估方式:全面性、多样性、赋权性)	EM: 0 F1 : 0.034 耗时: 4811 秒 Tokens: 177.23万	<ul style="list-style-type: none"> <li>通过rdf 五元组 (带类型) 抽取完成图谱构建</li> <li>问答阶段, 通过对query 中所包含实体、实体归属的概念实现Chunk 召回</li> <li>未利用语义、逻辑、符号等图谱技术栈</li> </ul>
OpenSPG-KAG (V0.5)	事实问答类任务+逻辑推理类任务 (评估方式:EM、F1)	EM: 0.625 F1 : 0.762 耗时: 4232 秒 Tokens: 227.6 万	<ul style="list-style-type: none"> <li>基于知识抽取、语义对齐、文本&amp;图互索引等完成图谱知识库构建</li> <li>基于逻辑符号引导的混合推理, 实现事实问答&amp;逻辑推理类任务。</li> <li>kag-model 小模型媲美大模型效果, 待开源...</li> <li>摘要、对话类任务, 待推进...</li> </ul>

<http://leaderboard.openkg.cn/>

即将推出

## 大模型知识增强系统评测

OpenKG旗下  
大模型利用外部知识库的能力评测



### 3大核心能力:检索、问答、推理

检索：从数据中检索信息；问答：根据数据回答问题；推理：需要综合外部库的信息，结合推理生成答案。



### 多种垂直领域: 医疗、金融、政务等

医疗知识问答、研报纠错、政务事项问答等。



### 多种外部知识库: 知识图谱、表格、图文、大图、半结构化数据等

本评测涉及不同形式的外部知识库上的不同任务，如知识图谱的链接预测和问答、表格问答、混合文本和表格问答等。

# 联系我们

KAG: <https://github.com/OpenSPG/KAG>

使用文档: [ReadMe](#)

(欢迎使用、反馈, 关注star)



(技术交流群)