



中国科学院计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

基于科学数据的检索增强生成 (RAG) 技术研究与应用

中国科学院计算机网络信息中心
韩芳

2024年9月20日

提纲

一、RAG技术与应用概况

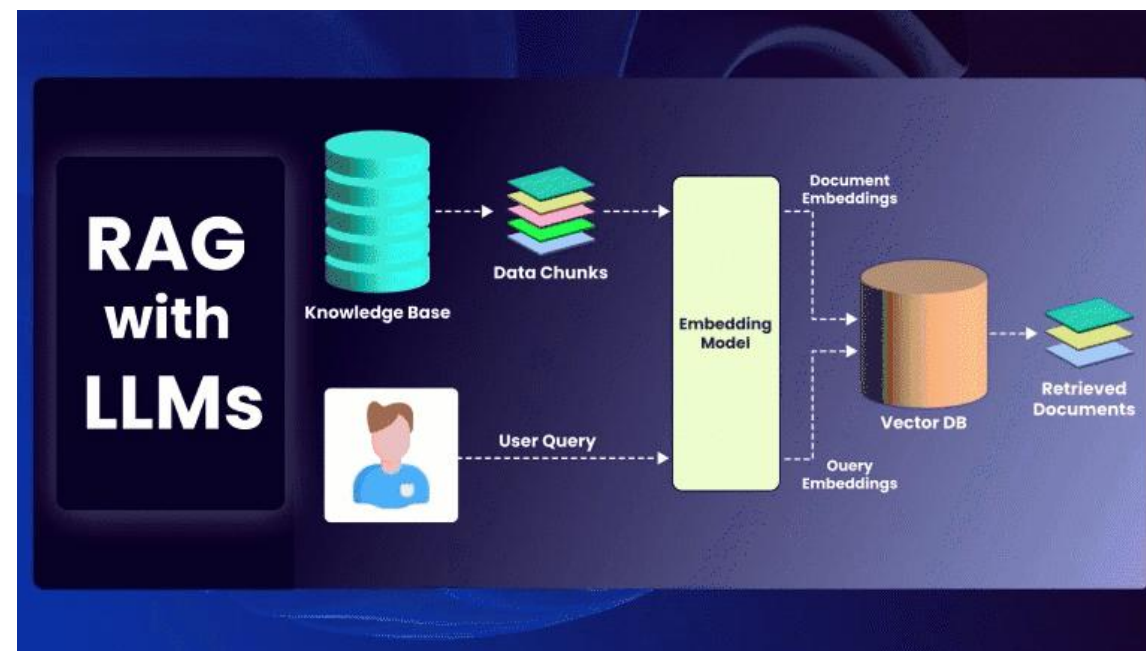
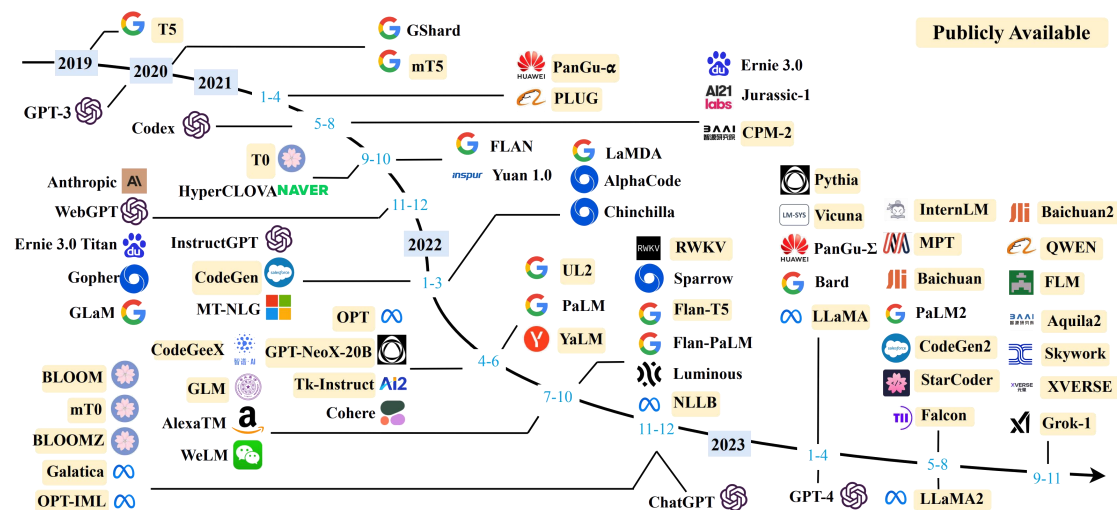
二、RAG核心技术问题研究

三、基于RAG的科学数据智能服务

基于RAG的智能服务

RAG (Retrieval-Augmented Generation, 检索增强生成) 是一种结合了信息检索和自然语言生成的技术, 主要应用于增强大型语言模型 (如GPT等) 的生成能力。

2020年由Meta AI提出



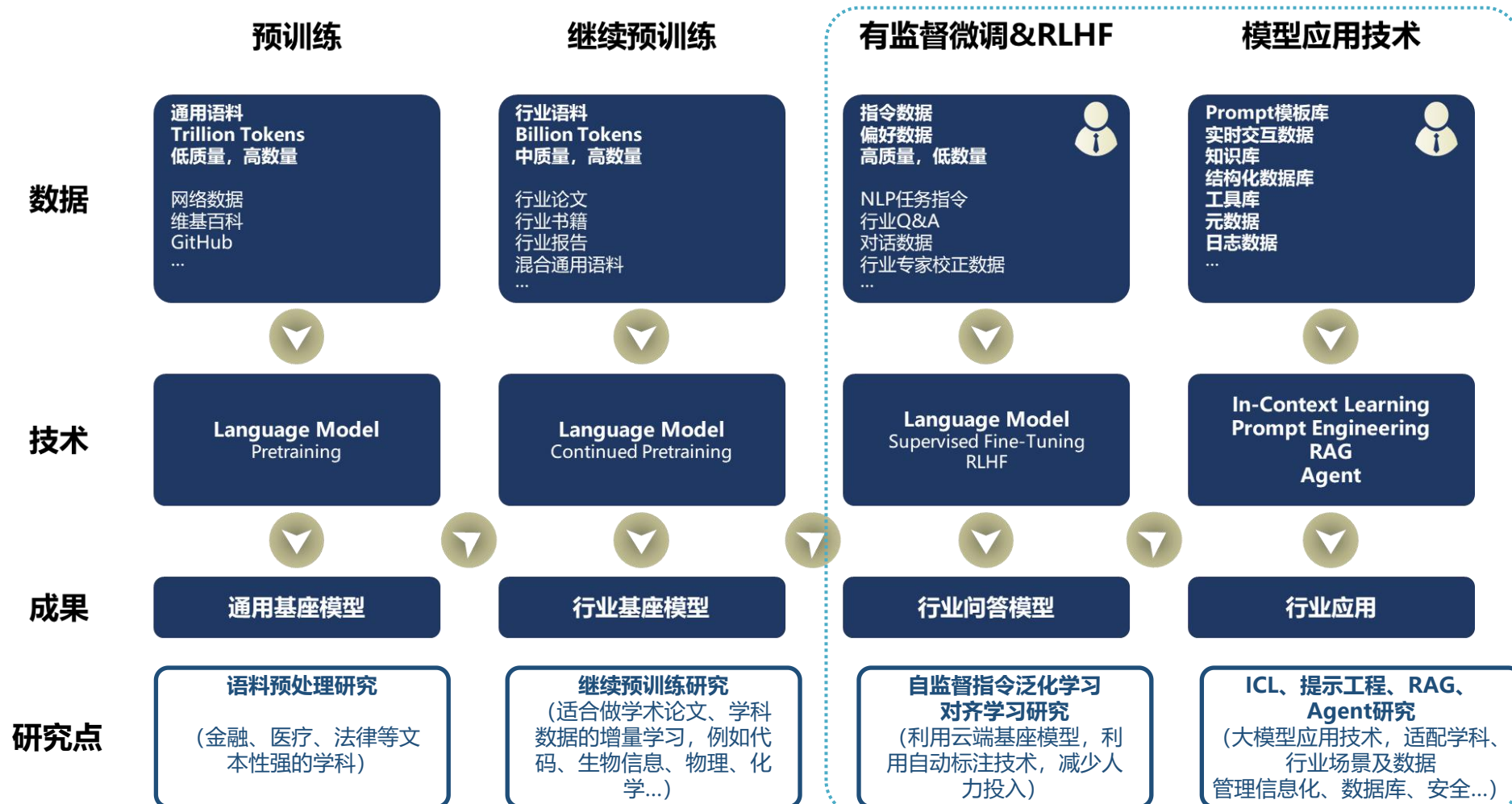
Zhao, Wayne Xin, et al. "A survey of large language models." arXiv preprint arXiv:2303.18223 (2023). <https://learnopencv.com/rag-with-llms/>

RAG-大模型应用技术

预训练阶段：模型在大规模通用数据集上学习基础知识和结构。

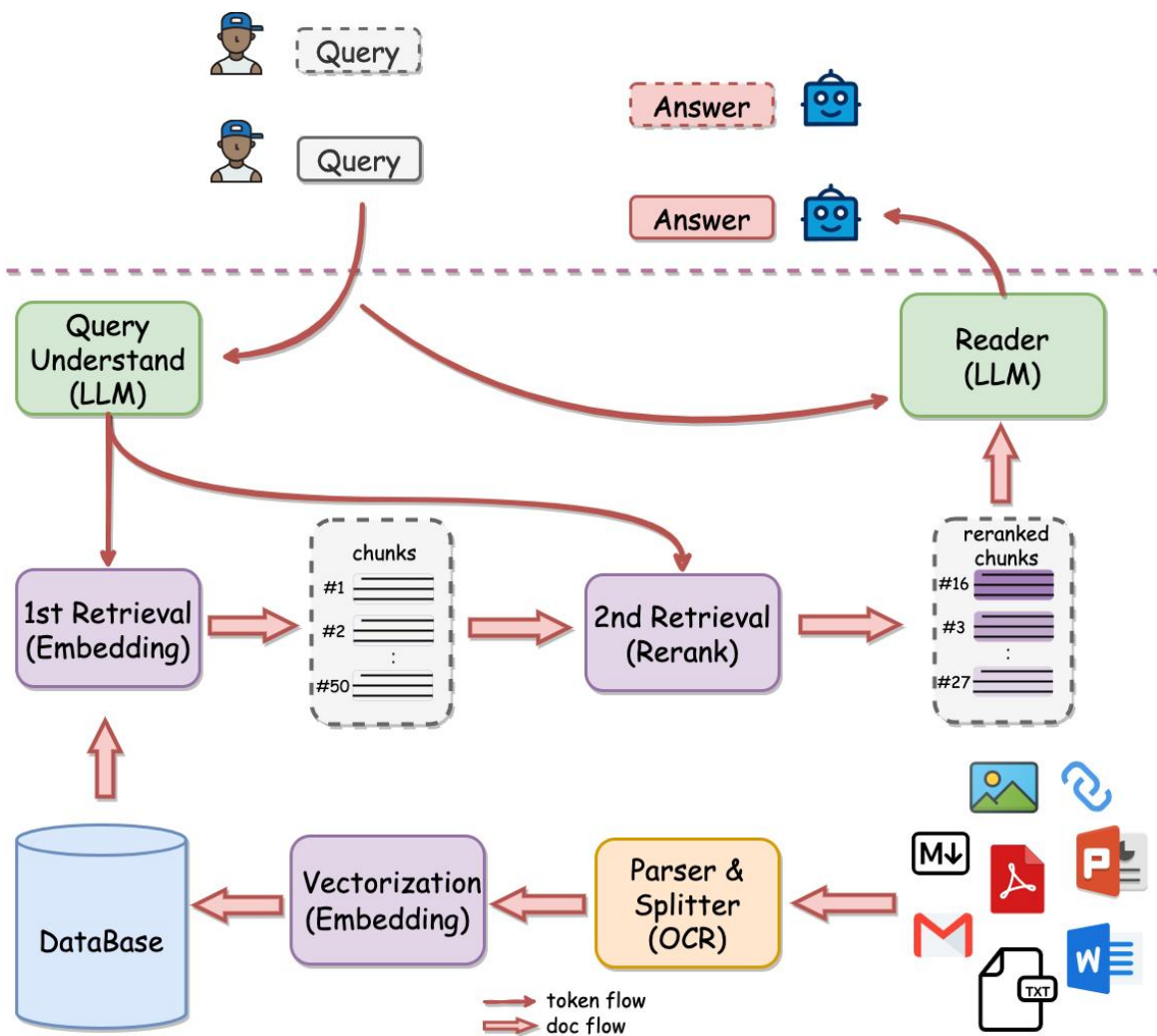
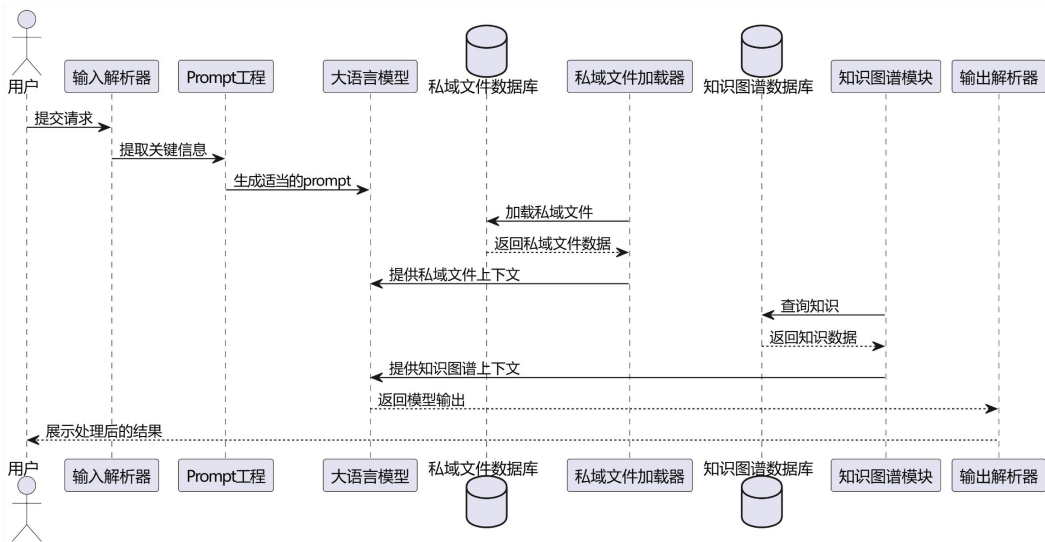
微调阶段：将预训练模型应用到特定任务的数据集上，通过再训练提升表现。

模型应用阶段：通过优化模型交互，使用 **RAG 技术**和 **Agent** 实现动态任务处理，并在各行业中应用。



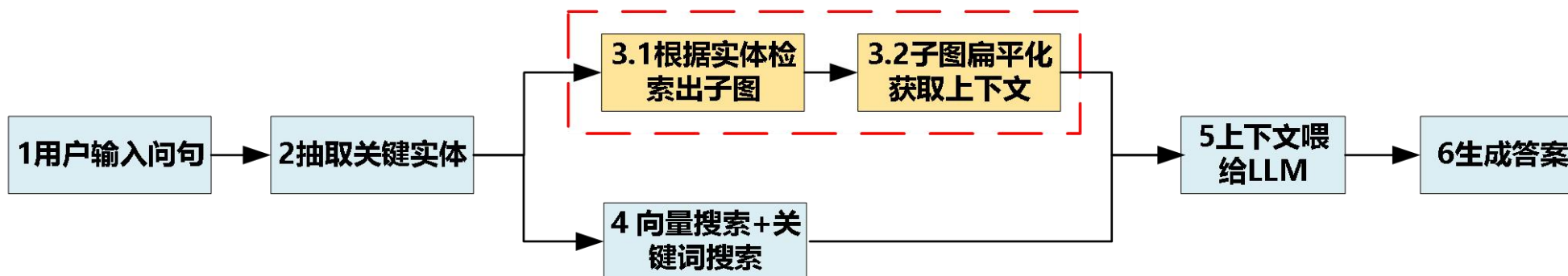
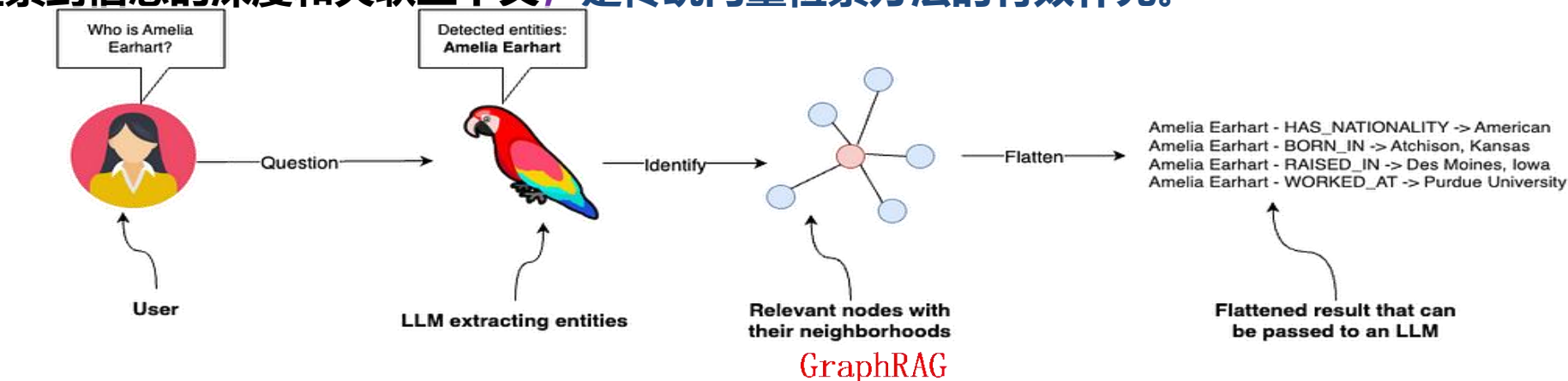
检索增强生成

- 输入解析器
- Prompt 工程
- 输出解析器
- 向量数据库
- 知识图谱模块
- 私域文件加载器
-



利用知识图谱提升RAG应用的准确性

图检索增强生成（GraphRAG）利用图数据库的结构化特性，以节点和关系的方式进行组织数据，增加了检索到信息的深度和关联上下文，是传统向量检索方法的有效补充。



Self-RAG (Self-Reflective RAG) 是一种增强的RAG范式，由华盛顿大学和IBM人工智能研究院的技术专家提出，核心思想是通过自省机制来优化生成任务的准确性和质量。

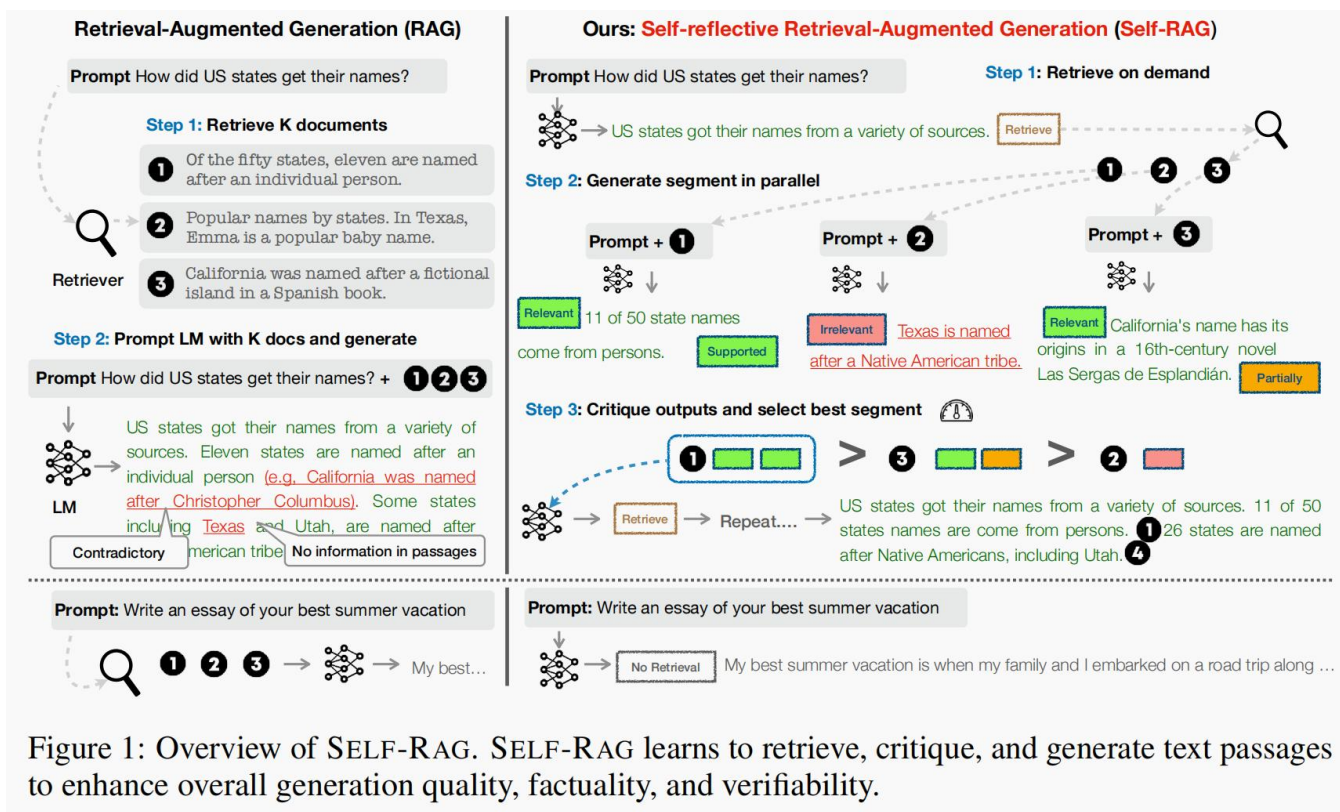


Figure 1: Overview of SELF-RAG. SELF-RAG learns to retrieve, critique, and generate text passages to enhance overall generation quality, factuality, and verifiability.

- 按需检索:** Self-RAG在生成内容时，首先判断是否需要检索外部知识。如果不需要，模型直接生成答案；如果需要，则进行检索。
- 自省标记:** 模型在生成过程中会插入自省标记（如[Retrieval]、[Relevant]等），这些标记帮助模型评估生成内容的相关性和支持度。
- 增强生成:** 在需要检索的情况下，Self-RAG会使用检索到的相关知识与输入问题一起生成多个响应，并通过评分算法选择最优答案。
- 评估指标:** Self-RAG设计了四种评估指标：检索判断（Retrieve）、知识相关性（IsRel）、响应支持度（IsSup）和响应有效性（IsUse）。

1. 客户支持

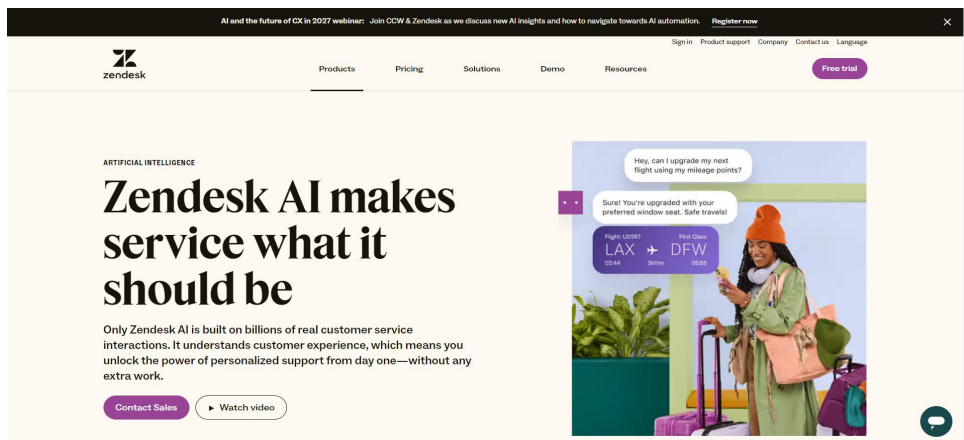
Zendesk AI

Zendesk AI 利用GPT-4及RAG技术增强客服自动化功能，提供更智能的聊天机器人和客户支持系统。

2. 医疗与健康

Mayo Clinic

Mayo Clinic 利用GPT-4为患者提供医疗信息查询和症状分析，帮助用户了解病情并获得健康建议。



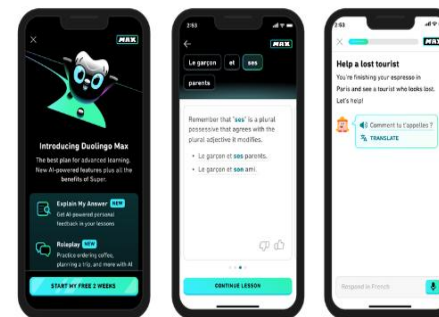
3. 教育

Duolingo Max

Duolingo Max是语言学习平台Duolingo的高级版本，集成了GPT-4，提供个性化的学习体验。该版本通过互动式练习和即时反馈帮助用户更快地掌握新语言。GPT-4可以生成动态的练习内容，回答用户提出的问题，并模拟真实对话场景

Khan Academy

Khanmigo是Khan Academy推出的一款虚拟教学助手，基于GPT-4，为学生和老师提供支持。它能够解释复杂的概念，提供问题解答，并通过互动式的学习方法帮助学生理解各种学科内容。对于老师来说，Khanmigo可以生成教学资源 and 个性化的教学建议。



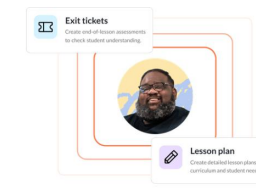
On-demand AI-powered support for education.

For teachers

Experience the best AI for teachers.

Built for educators by educators, with your needs in mind. Khanmigo simplifies your workflow while keeping your work and your student data private and secure.

Sign up for free Learn more



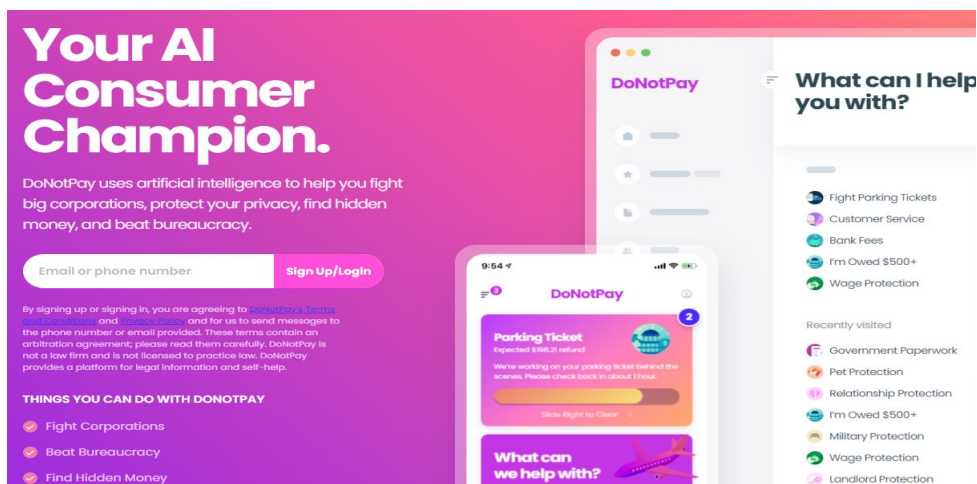
1. 法律

DoNotPay

DoNotPay 使用GPT-4生成法律文件、提供法律建议，帮助用户解决小额纠纷、申诉罚单等。

Harvey AI

Harvey 利用LLM为法律事务所提供法律研究、合同审查和文档生成服务，提升法律工作的效率。



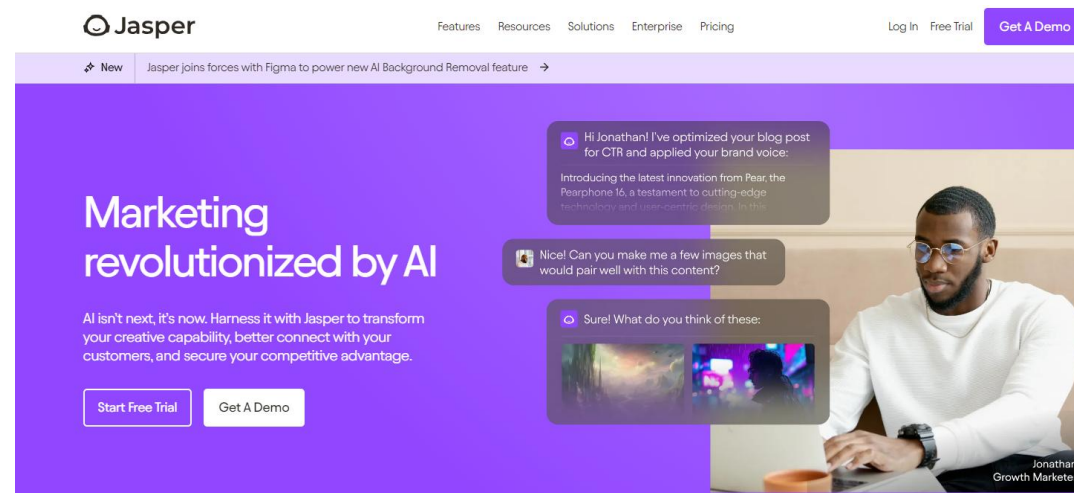
2. 文案生成与写作

Jasper AI

Jasper AI 基于LLM提供内容生成服务，包括文章、广告文案、博客内容等，帮助创作者提高写作效率。无论是市场营销专业人员还是内容创作者，Jasper AI都能提供定制化的写作建议。

Copy.ai

Copy.ai 使用GPT-4生成市场营销文案、社交媒体帖子和电子邮件，帮助企业和个人更快速地创建内容。该工具帮助用户快速创作吸引眼球的内容，适用于多种营销场景。



提纲

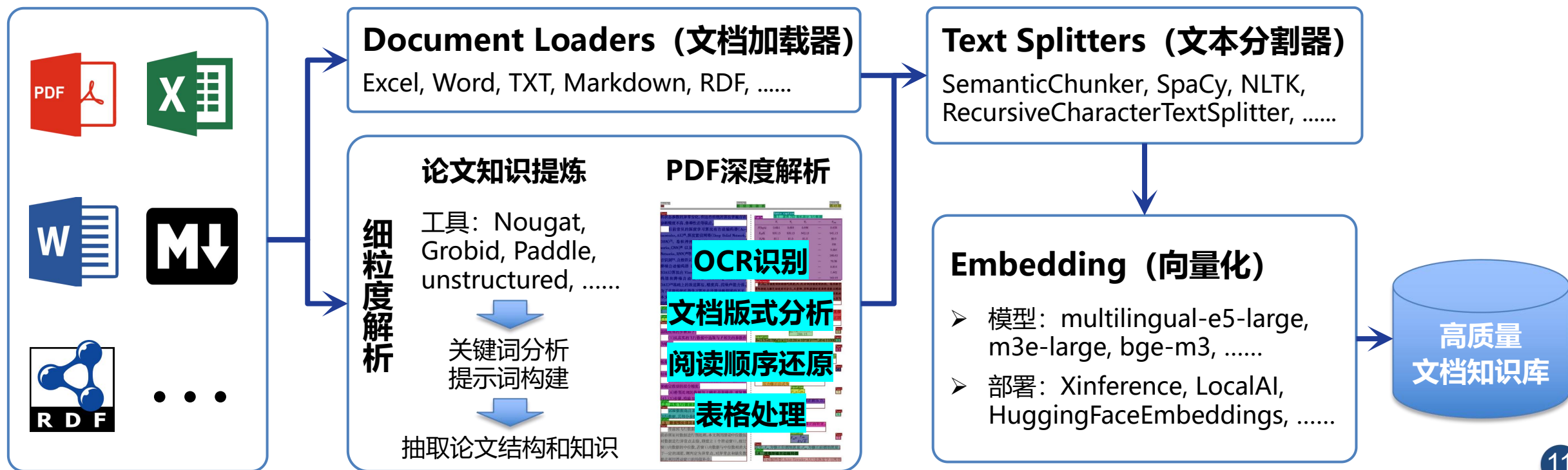
一、RAG技术与应用

二、RAG核心技术问题研究

三、基于RAG的科学数据智能服务

技术难点：文档类型的多样性和版式的复杂性，会造成**文档结构和内容解析不够准确、分块过程中易丢失语义信息**等问题，进而影响**RAG的检索质量和回答准确性**。

解决方案：引入**多元化的文档加载器**，实现多类型文档内容的全面解析；采用更**细粒度**的方式实现**论文知识提炼**和**PDF深度解析**，有效提升文档解析的准确性。通过**文本语义分割器**对解析结果进行精细化的分块处理，并结合**向量化技术**，实现**高质量文档知识库**的智能构建，显著提升RAG的检索质量和回答准确性。



技术难点：用户的多样化提问、复杂的知识库以及庞大的数据量，使得海量知识库中数据检索的准确性成为了一项重要挑战，传统的检索方法难以保证检索结果的准确性与相关性。

解决方案：通过问句分类、问题重写、语义扩展、思维链等手段，实现问句语义的精准解析。采用多路召回策略，整合向量检索、全文检索、业务数据库查询、工具调用等多种方式，全面增强知识召回的广度和深度。并通过检索重排序技术，精准发现最相关的文本片段，显著提升检索结果的准确性与相关性。



上下文窗口高效管理

技术难点：在上下文窗口管理中，如何增强LLM对上下文理解的准确性是一项挑战。随着对话轮次的增长，上下文窗口大小可能超出模型的Token限制，且随着对话窗口数量的增多，系统难以有效存储大量的对话数据。

解决方案：构建高质量的提示词（Prompt），引导模型能够更准确地理解上下文；采用多种长上下文压缩策略，如分块并行总结、分块串行总结和分块重排优选等，压缩对话历史，扩展上下文窗口容量。同时，通过合理的缓存淘汰算法及数据存储策略，实现大量对话数据的有效存储，确保上下文窗口高效管理。

提示词 (Prompt) 要素

关键要素	角色、指令、上下文、限制条件、输入数据、输出指示、示例
辅助要素	简洁性、语气、强调、重复检查、多样化输出、思维链提示、上下文持久性、.....



缓存淘汰算法

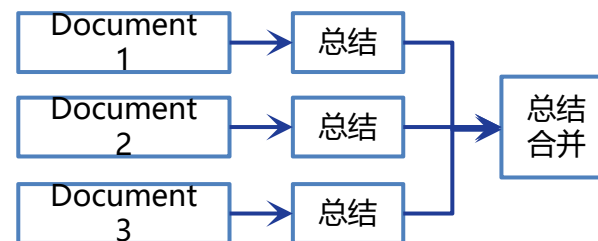
- LFU: 最不经常使用算法
- LRU: 最近最少使用算法
- ARC: 自适应缓存替换算法
- FIFO: 先进先出算法
-

长上下文压缩策略

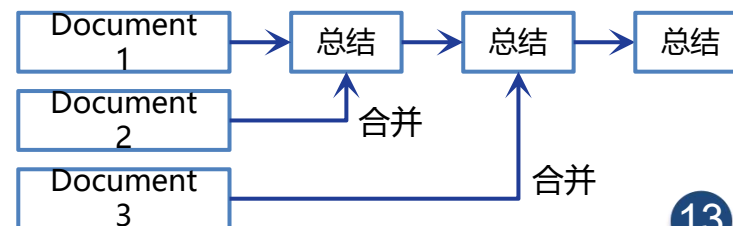
一次性总结 (Stuff)



分块并行总结 (Map Reduce)



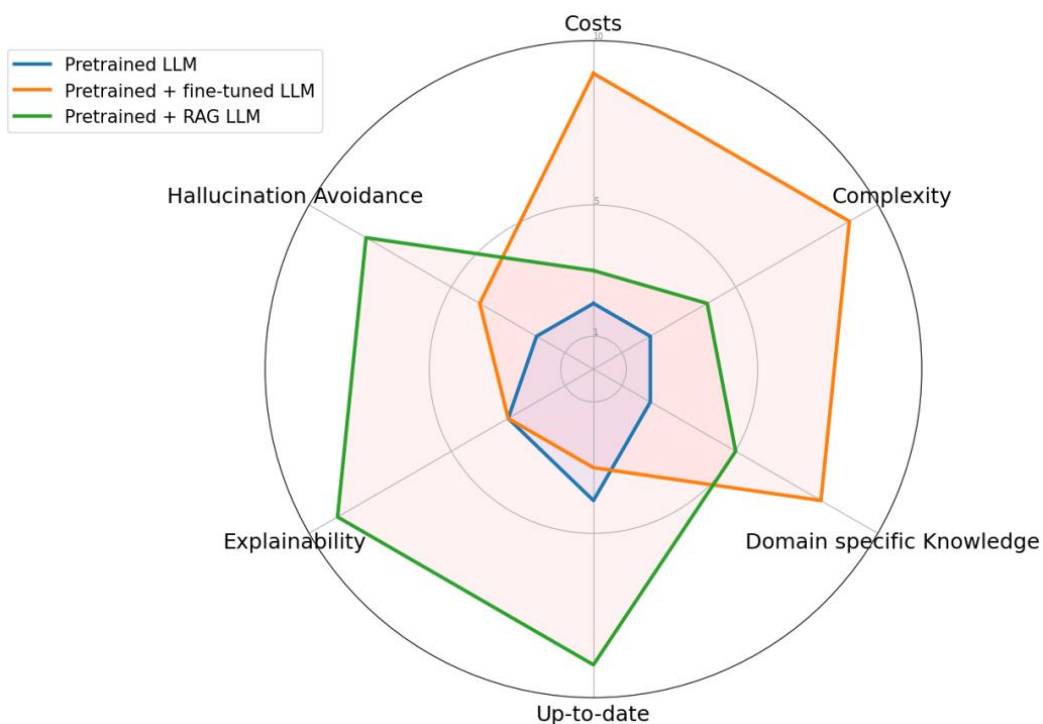
分块串行总结 (Refine)



模型能力增强与优化

技术难点：在检索增强生成（RAG）中，**对大语言模型能力的依赖程度较高**，且模型行为难以针对特定任务进行精确调整。同时，RAG在**性能方面**也存在一定的局限性。

解决方案：采用“**RAG+微调**”相结合的策略，融合两种方法的优势。利用RAG进行动态数据集成，提升**回答的准确性和可解释性**；同时，通过微调技术进一步增强**模型的能力和整体性能**，以适应特定任务需求。



指标	RAG	微调
动态数据	√	×
数据要求	√	×
准确性	√	×
可解释性	√	×
性能	×	√
多语言	×	√
能力定制	×	√

全量微调	高效微调
模型的所有参数都参与训练，以适应特定任务或领域的数据。这种方法需要较大的计算资源和时间，但能够充分利用模型的全部容量来优化性能。	仅调整模型的一部分参数或引入额外的少量参数来适应特定任务或领域的的数据，可以显著减少计算资源和时间需求。代表性算法如：LoRA、QLoRA、P-Tuning等。

提纲

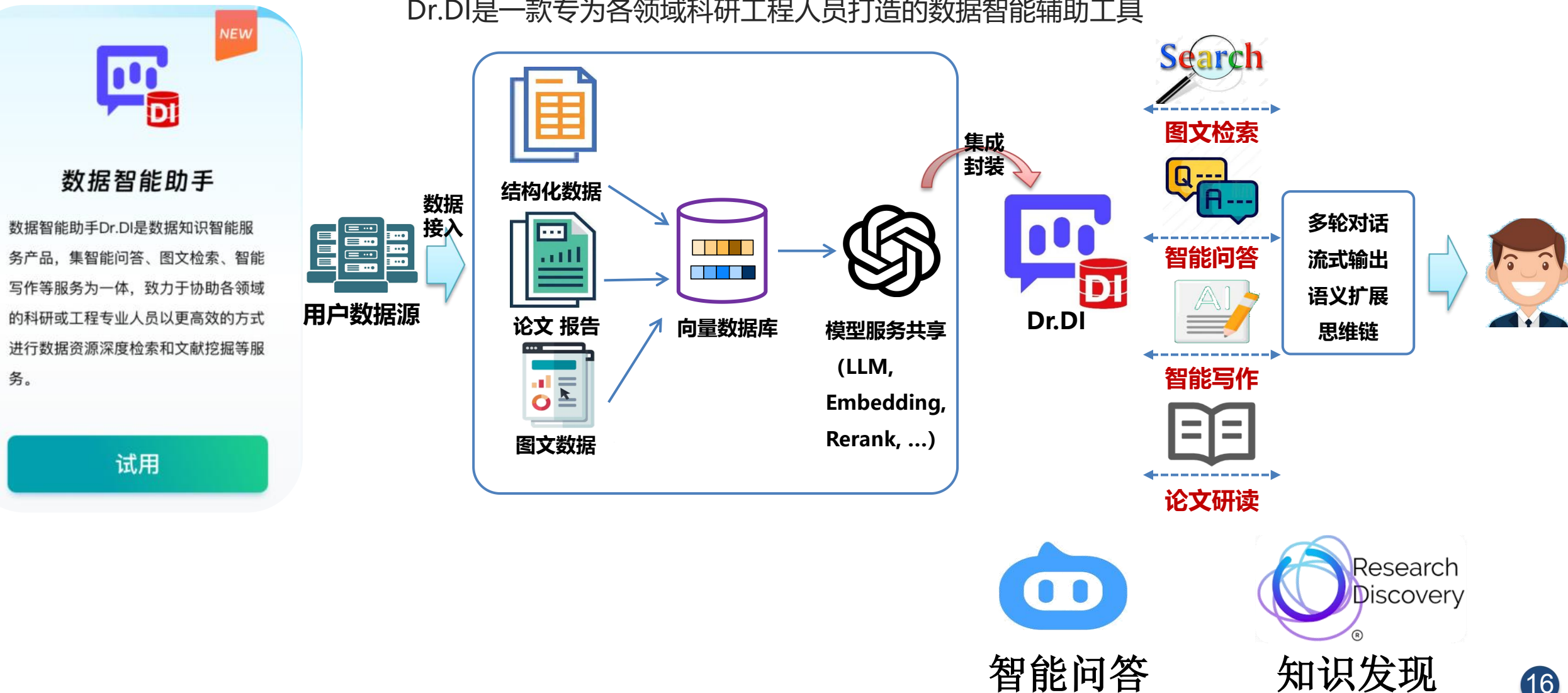
一、RAG技术与应用

二、RAG核心技术问题研究

三、基于RAG的科学数据智能服务

数据智能助手-Dr.DI

Dr.DI是一款专为各领域科研工程人员打造的数据智能辅助工具



试用

GPU服务器一台

- GPU：显存 $\geq 24\text{GB}$ ，推理性能 $\geq \text{RTX A6000}$ ，型号为NVIDIA；若CPU部署需使用在线模型API
- 推荐内存64G，CPU16核，硬盘2T，视本地数据量而定
- 操作系统：Ubuntu 22.04 LTS
- 大语言模型（开源）：Qwen2-7B-Instruct
- Embedding模型（开源）：multilingual-e5-large
- 图文多模态模型（开源）：chinese-clip-vit-large-patch14
- 数据库（开源）：Qdrant、MongoDB
- 开发语言：Python 3.10
- 开发框架：LangChain、Flask、Xinference 等
- 其他软件环境：Docker、nvidia-container-runtime 等



NVIDIA®



mongoDB



Qwen



LangChain

请介绍一下地不容

Search

Try: 请介绍一下地不容

Try: 动车气动性能是怎么评价的?

Try: 动车组车内空调通风系统测试需要试验哪些内容?

Try: 地铁的通风系统评测依据是什么?

Try: 木棉和轻木有哪些区别?

Try: azylidynemethylum的化学式和CAS号

地不容 (*Stephania epigaea* H. S. Lo) 是防己科千金藤属的多年生草质藤本植物，全株无毛；块根较大，呈扁球形；茎稍粗，枝、叶折断有紫红色汁液渗出；单叶互生，叶片薄，三角状扁圆形；花小，暗红色，花瓣三枚，单性，雌雄异株；果梗短而肉质，核果红色，果核倒卵圆形；花期5—6月；果期6—9月。[1] [2] [3] [4] [5] [6] [7] 地不容分布于云南、四川，生于山地灌木丛中，林缘或石缝等海拔2000米处。[2] [7] [8] [9] 地不容喜阴湿，耐阴，耐旱。[2] [7] 地不容采用播种繁殖。地不容味苦，性寒，有小毒，有清热解毒、镇静、理气、止痛的功效，可用于解疮毒，主治疟疾，食积腹痛。[2] [4] [6] [7] 地不容叶翠绿，盆栽适合阳台、庭院等处，观赏性佳

参考依据:

[1] ☆ 地不容 2 播报 编辑 讨论 口上视频 防己科千金藤属植物一展开2个同名词条√ 地不容 (*Stephania* 稍粗，枝、叶折断有紫红色汁液渗出；单叶互生，叶片薄，三角状扁圆形；花小，暗红色，花瓣三枚，单性，雌雄异株；果梗短而肉质，核果红色，果核倒卵圆形；花期5—6月；果期6—9月。[8] 地不容分布于云南、四川，生于山地灌木丛中，林缘或石缝等海拔2000米处。[地不容喜阴湿，耐阴，耐旱。地不容采用 播种繁殖。地不容味苦，性寒，有小毒，有清热...

引用: 地不容-百科.pdf 下载

[2] 植物名称: 地不容 & 科: 防己科 & 属: 千金藤属 & 分布地: 云南、四川

引用: 地不容植物.csv 下载

[3] 地不容 (*Stephania epigaea* H. S. Lo) 是防己科千金藤属的多年生草质藤本植物，全株无毛；块根较大，呈扁球形；茎稍粗，枝、叶折断有紫红色汁液渗出；单叶互生，叶片薄，三角状扁圆形；花小，暗红色，花瓣三枚，单性，雌雄异株；果梗短而肉质，核果红色，果核倒卵圆形；花期5—6月；果期6—9月。

引用: 地不容-介绍.pdf 下载

[4] 地不容 形态特征 草质落叶藤本，全株无毛，有硕大的扁球状块根，暗灰褐色，嫩枝稍肉质，紫红色，有白霜。叶扁圆形，稀近圆形，长3-5厘米，宽5-6.5厘米，下面稍粉白，掌状脉3块以上。叶柄长4-6厘米，质状着生叶片基部约1-2厘米处。花序梗长1-4厘米，丛生几个至十几个小聚伞花序，每序2-3花，花小，紫色，核果红色，倒卵形，背部两侧有小横肋16-20条。块根含多量甙法丹质和轮黄藤宁，能清热解暑、镇静、理气止痛。

引用: 地不容-形态特征.md 下载

[5] 植物名称: 海南地不容 形态特征: 多年生藤本。块根球形或不规则球形，露于地面。老茎稍木质化，枝粗，有直沟槽，枝、叶含淡黄色或白色乳汁，全株无毛。叶互生；叶柄短，通常与叶片近等长或稍短，盾状着生；叶片三角状圆形，长和宽均10-16cm，有时较小，先端短渐尖，基部圆或近平截，有时微凹，边缘浅波状，或呈具星状粗齿，有时近全缘；掌状脉5条，网状脉具有小乳突，薄纸质。

引用: 海南地不容.pptx 下载

[6] 【中药名】地不容diburong 【别名】金不换、一点血、山乌龟。【英文名】Radix 【来源】防己科植物血散薯*Stephaniadielsiana*Y.C.Wu.或地不容(山乌龟)*Stephania* 【植物形态】多年生草质藤本，全株无毛。块根近长圆纺锤形或不规则块状，常露于地面，外皮褐色，粗糙，有许多疣状小乳突。茎常带紫红色，枝、叶折断有红色乳汁流出。叶互生，纸质；盾状着生；叶片阔三角状卵形，先端有小突尖，基部截形或近圆形，全缘。聚伞花序腋生，雌雄异株；雄...

引用: 地不容-中药信息.jpeg 下载

[7] 海南地不容 ①播报 编辑 上传视频 中药展开2个同名词条√ 同义词条由国家中医药管理局中药名词术语成果转化与规范推广项目提供内容， 声明温馨提示: 保护野生植物，维护生态平衡。禁止非法采集、买卖、收购、加工国家保护野生植物或破坏其生长环境 海南地不容，中药名。为防己科植物海南地不容*Stephaniahainanensis*H.S.LoetY.Tsoong的块根。分布于海南省白沙、琼中、儋县等地。具有止痛，消肿解毒之功效。常用于胃痛，外伤疼痛，疮疗脓肿

引用: 海南地不容-百科.png 下载

[8] 地不容，在云南除滇西南外，其他各地均有分布；也分布于贵州的罗甸、安龙、贵定；广西、湖南、广东、湖北、四川等地也有分布。

引用: 地不容-分布地.txt 下载

[9] 地不容，生于海拔600~1000m的山谷溪边、林中、石缝及峭壁上。常生长于石山，也常见栽培。在云南除滇西南外，其他各地均有分布。全世界约有50种，中国有30种，云南有18种。山乌龟在巧家、永善、镇雄、彝良等县的岩壑又深箐中，都有生长。

引用: 地不容-简介.docx 下载

Dr.DI结合检索增强生成技术，通过从大量文档中检索相关信息，生成高质量、上下文相关的答案。

增加了OCR等文本识别能力，扩展了常见文本类格式的识别和问答，目前已支持文件格式包括：pdf、excel、csv、txt、docx、md、pptx、jpg、png、jpeg

支持多中心分布式资源语义检索

目前支持4类问句：实体找属性、属性找实体、2D相似、3D相似，其中2D、3D均为实时查询pubchem。

温度+状态

问句样例：常温下呈固态的化合物有哪些？

所需条件：

- 温度：常温/高温。注：常温为25°C，转化为开尔文温度为25+273.15=298.15K。
- 状态：固态/液态。注：高温为100°C，转化为开尔文温度为100+273.15=373.15K。

分析结果：

- 固态：熔点 (T.Tm) >= 开尔文温度 (298.15K/373.15K)
- 固态：液态 (T.Tm) <= 开尔文温度 (298.15K/373.15K)

溶解性

问句样例：哪些化合物易溶于水？

所需条件：

- 溶解性：易溶/可溶/微溶/难溶

chemdb库里的值需要做转化：将库中查到的数值，先求反对数，再乘分子量，再除10。具体公式为：

$$-\text{LOG}_{10}(\text{M.Molecular_Solubility}) * \text{B.MolWeight} / 10$$

分析公式结果位于哪个区间：

- 易溶：公式结果 >= 10
- 可溶：公式结果 >= 1 AND 公式结果 <= 10
- 微溶：公式结果 >= 0.01 AND 公式结果 <= 1
- 难溶（或不溶）：公式结果 <= 0.01

摩尔质量在10-100之间、常温下呈固态、可溶于水的化合物有哪些？

Search

Try: water的化学式是什么？

Try: 化学式为H2O的化合物有哪些？

Try: CAS号为7732-18-5的化合物是什么？

Try: 摩尔质量在43-45之间、可溶于水的化合物有哪些？

Try: 分子量在15-20之间、常温下呈液态的化合物有哪些？

Try: 摩尔质量在10-100之间、常温下呈固态、可溶于水的化合物有哪些？

Try: water和hydrogen peroxide有哪些异同点？

Try: water的2D相似化合物有哪些？

Try: hydrogen peroxide的3D相似化合物有哪些？

√ 已识别检索意图：根据条件检索化合物

√ 已抽取检索条件：mol_weight_max: 100, mol_weight_min: 10, molecular_solubility: 可溶, state: 固态, temperature_condition: 常温

√ 答案总结：

1. Ethyllithium (C2H5Li), 相对摩尔质量: 36.0。
2. Methylboronic acid (CH5BO2), 相对摩尔质量: 60.0。
3. Urea (CH4N2O), 相对摩尔质量: 60.0。
4. 2-Ammonioacetate (C2H5NO2), 相对摩尔质量: 75.0。
5. 2-Hydroxyacetate (C2H3O3), 相对摩尔质量: 75.0。
6. Methyl carbamate (C2H5NO2), 相对摩尔质量: 75.0。
7. Dimethylamine hydrochloride (C2H8ClN), 相对摩尔质量: 82.0。
8. Ethanamine hydrochloride (C2H8ClN), 相对摩尔质量: 82.0。
9. Amino(cyanamido)methaniminium (C2H5N4), 相对摩尔质量: 85.0。
10. N'-Formylformohydrazide (C2H4N2O2), 相对摩尔质量: 88.0。
11. (R)-2-Ammoniopropanoate (C3H7NO2), 相对摩尔质量: 89.0。
12. 1-Hydroxy-2-methylpropan-2-aminium (C4H12NO), 相对摩尔质量: 90.0。

问：摩尔质量在10-100之间、常温下呈固态、可溶于水的化合物有哪些？

化合物名称	化学式	相对摩尔质量	CAS号
ethylithium	C2H5Li	36	811-49-4
methylboronic acid	CH5BO2	60	13061-96-6
urea	CH4N2O	60	57-13-6
2-ammonioacetate	C2H5NO2	75	56-40-6
2-hydroxyacetate	C2H3O3	75	79-14-1





以文搜图：用户输入一段描述性的文本，系统通过分析文本内容，提取其特征向量，并在数据库中搜索与之匹配的图像。本次升级增加了基于图片内容的查询功能，如：“红黄相间的花朵”。

以文搜图 | 以图搜图





检索文本:

检索示例: 丛尾草 大花茄 宁夏枸杞 少花龙葵 巨嘴鸟 报春花 白英 地海椒 曼陀罗 白辛树 矮牵牛 粉色风铃野茉莉 茉莉果 龙葵 银钟花 香机 底盘 民航发动机 直升机旋翼 猛禽战斗机 驱逐舰 导弹发射 苏35战斗机

按图片文字说明检索结果:

 名称: 小花02	 名称: 小花01	 名称: 小尾花	 名称: 半日花03
--	--	---	--

按图片内容检索结果:

 名称: 矮牵牛01	 名称: 舞春花01	 名称: 矮牵牛02	 名称: 报春花01
--	--	--	---

以图搜图：用户上传一张图片，系统提取该图片的特征向量，在数据库中搜索相似的图像。例如，上传一张特定型号的战斗机，系统会返回类似型号的飞机。





智能问答 | 图文搜索 | 智能写作 | 思维链

以文搜图 | **以图搜图**

上传图片文件

输入图片URL

以图搜图检索结果:

 名称: 少花龙葵05 相似度: 0.846806	 名称: 龙葵04 相似度: 0.843497	 名称: 少花龙葵03 相似度: 0.831888	 名称: 龙葵01 相似度: 0.831611
--	--	--	--

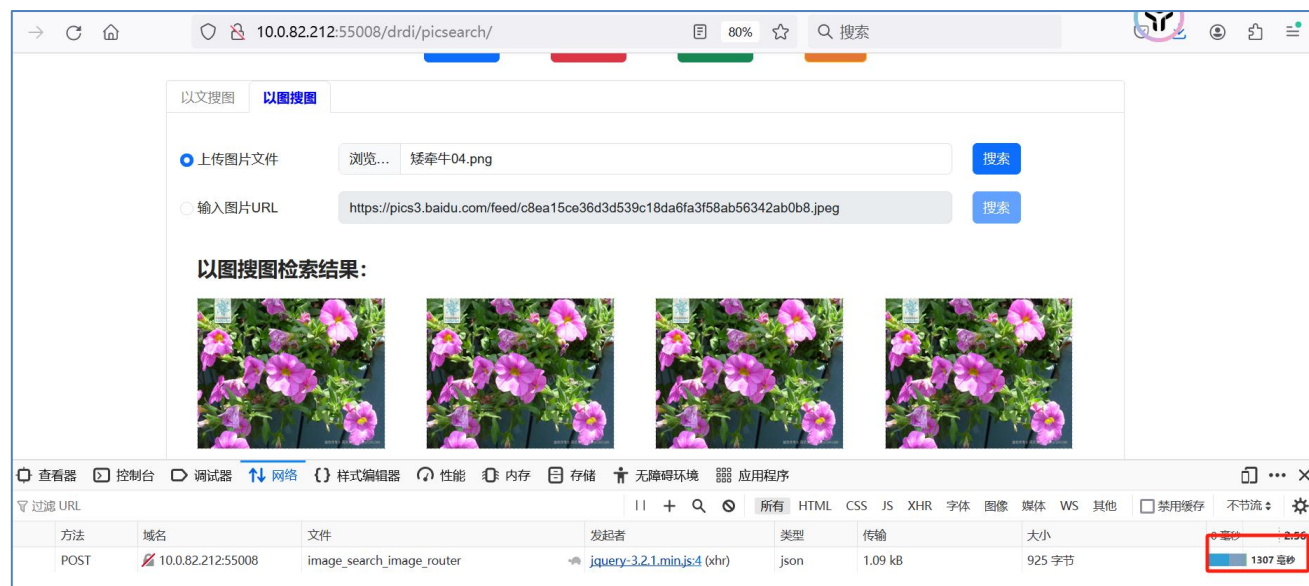
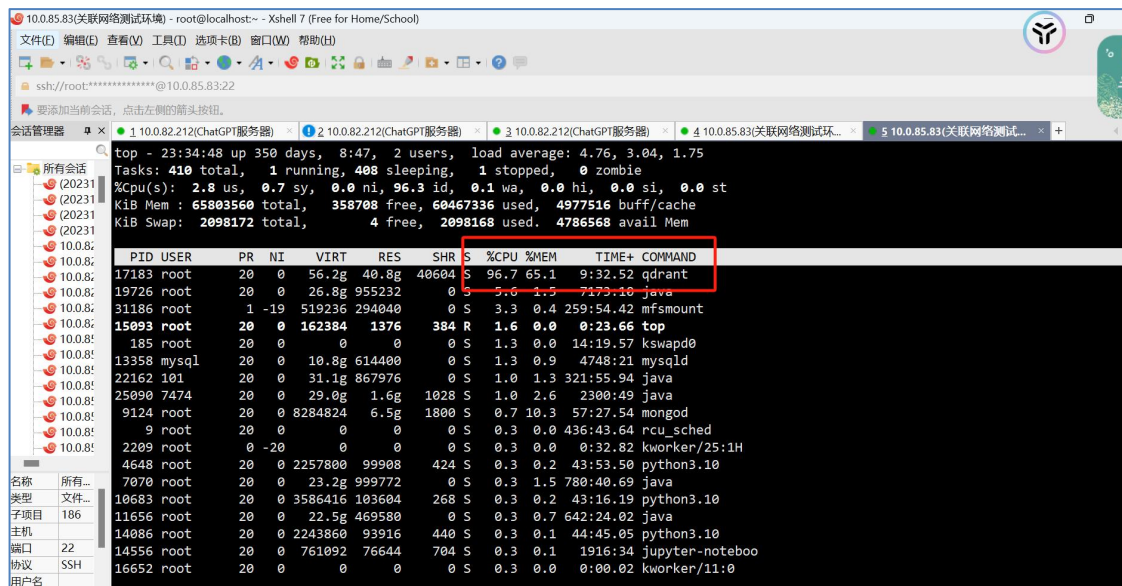
Dr.DI-图文检索 性能测试

针对图文检索模块进行大数据量下的性能测试：

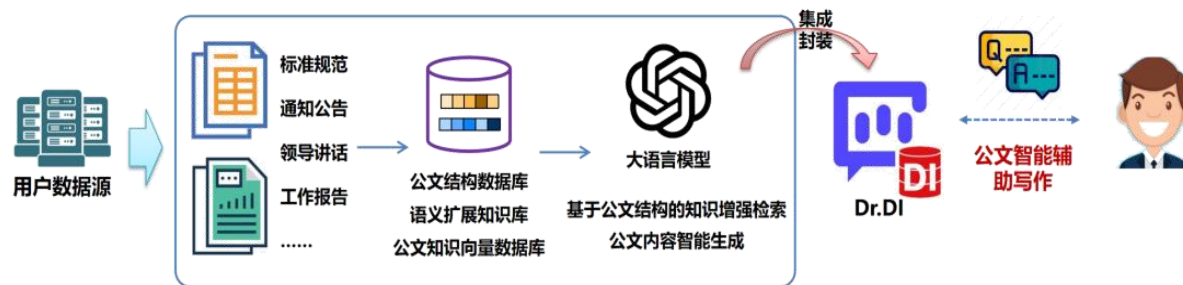
- 千万级图文数据量下，图文检索性能达到**秒级**。
- 千万级图文数据量下，图文检索需要占用较高的内存；入库时间较长，但检索时间较短。

详细测试信息：

- 数据存储：1000万个200K的图片，存储空间2T；向量数据库60G；内存70G；开10个多线程，A100机器完整处理约20小时。
- 查询时间：百万级图文检索平均查询时间约200ms；千万级图文检索查询时间约800~2500ms。



基于用户的公文写作需求，结合**本地公文**知识，采用基于公文结构的**知识增强**检索、公文内容**智能生成**关键技术，借助大模型技术理解公文结构、生成和整合公文内容，实现基于公文知识的公文智能辅助写作。



题目	2020年邢台市烟草局工作总结和2021年工作计划
<input checked="" type="checkbox"/> 生产经营	稳定经济
<input type="checkbox"/> 营销与服务	输入营销与服务主题，例如：配套制定专项规划，**开展经济责任风险自查自纠梳理风险点。
<input checked="" type="checkbox"/> 市场监管	常态化综合治理
<input type="checkbox"/> 基础管理	输入基础管理主题，例如：强化干部队伍建设，完善绩效考评。
<input type="checkbox"/> 人员队伍	输入人员队伍主题，例如：培养年轻干部，提高队伍能力素质。

生成文章

2020年邢台市烟草局工作总结和2021年工作计划

本年主要工作情况

一、生产经营

根据提供的资料，我对“生产经营”方面的工作成果进行总结如下：

在疫情防控和生产经营两手抓的政策指导下，全区卷烟销量持续增长，1-11月同比增长1.76%，税金同比增长3.04%。

经济效益稳中有升，1-11月税利同比增长3.45%，市场状态企稳向好，卷烟动销平稳，品牌培育效果明显。

坚持“总量控制、稍紧平衡”调控方针，海口市局全年经济运行保持平稳发展，卷烟销量稳定增长，经济效益持续增长，市场状态企稳向好，自有品牌销量同比增长11.50%。

牢牢把握“十六字”方针，全年片区卷烟销量同比增长1.5%，自有品牌销量同比增长10.28%，网络建设不断完善，深化零售终端建设，持续推进诚信互助小组建设。

抓住关键点，卷烟销售量价齐增，强调研、抓时点、保货源，明确培育重点，做好品规进出管理，找准管理痛点，营销基础逐步提升，瞄准客户焦点，网建工作立足市场，加强客户服务力度，推动零售终端建设，实施客户分类指导。

二、市场监管

市场监管工作成果总结：1、持续完善高效便民政务服务体系，落实“放管服”改革要求，推广“一网通办”，建立政务服务“好差评”制度；2、聚焦卷烟市场治理，开展行动维护市场秩序，深化“双随机、一公开”与APCD工作法融合，筑牢市场监管防线；3、深入推进诚信建设，巩固互助小组建设成果，制定评价标准与考评措施，将“党建+”与互助小组工作深度融合；4、强化人员学习培训，提升综合素质，开展作风整顿，全面建设纪律组织保障体系；5、查处涉烟案件415起，查获非法卷烟466件，市场净化率达96.5%以上，实现绩效管理新突破。

总结：通过持续深化市场监管，加强打破假网工作，推动专卖管理效能升级，提升市场净化率和绩效管理水，全面提升市场监管效能，有效维护卷烟市场秩序，为市场规范经营和社会治理作出积极贡献。

明年工作计划

一、生产经营

工作计划：

生产经营方面的工作将从以下几个方面展开：

1、强化稳的状态，推动经济运行再上新台阶

根据疫情发展态势，科学精准优化防控措施，保障市场供应，提振客户经营信心；

强化物流现场管理，提高运行质量，实现高效率、低故障运行；

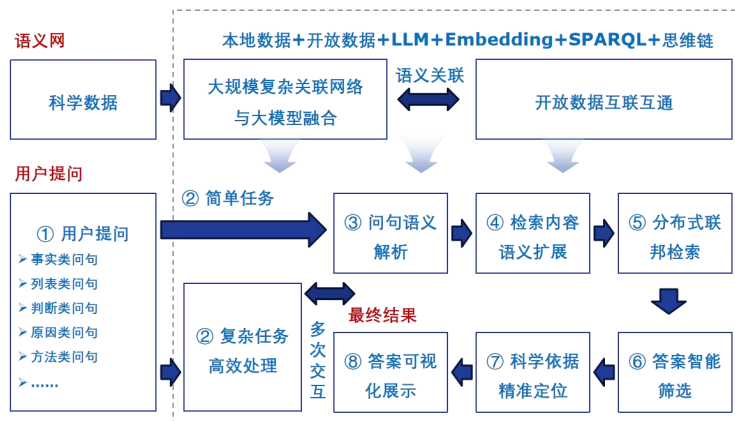
强化物流成本精细化管理，提升核心业务自主管控水平。

2、全面加强质量管理体系建设

细化梳理程序文件制定标准，优化管理流程；

加强质量管理体系与精益课题、QC小组活动的对接，延伸至基层：

针对用户提出自然语言**复杂问句**，系统运用**思维链技术**，依托大语言模型强大的**Function Call**能力，智能拆解用户提出的复杂问句并依次调用本地工具进行处理。同时，页面会**动态展示推理和执行过程**。在所有行动完成后，系统将自动生成**总结性的答案**。Dr.DI的思维链功能可以大幅提升科研人员处理复杂问题的效率，加速科研工作的进程。



科研助手Dr.DI - 思维链

智能问答

图文搜索

智能写作

思维链

用于治疗皮炎的植物有哪些？这些植物有什么用途？它们的化学成分是什么？

Search

点击查看工具列表

工具名称	工具描述
query_disease_tool	查询治疗某种疾病的植物
query_property_tool	查询某个植物的某个具体属性
query_chem_tool	查询某个植物的化学成分
query_color_tool	根据省份、花期、花色查询植物
design_planting_plan_tool	整理植物的种植方案

Try: 用于治疗皮炎的植物有哪些？这些植物有什么用途？它们的化学成分是什么？

Try: 在云南省内步行路两侧种植两种花，一侧在5月份开花，花色白色，一侧在10月份开花，花色黄色，请问有哪些种植方案？



Action: tool=query_disease_tool, tool_input={'jibing': '皮炎'}

Observation: ['流苏子', '博落木']

Action: tool=query_property_tool, tool_input={'plant': '流苏子', 'pLabel': '用途'}

Observation: {'result': '流苏子的用途是：根辛辣，可治皮炎。'}

Action: tool=query_chem_tool, tool_input={'plant': '流苏子'}

Observation: {'result': '流苏子主要含有黄酮类化合物，其中一些化合物具有抗肿瘤、抗氧化和抗炎等药理活性。'}

Action: tool=query_property_tool, tool_input={'plant': '博落木', 'pLabel': '用途'}

Observation: {'result': '博落木的用途是：该植物外用于治疗皮疹，慢性眼炎和去除疣。'}

Action: tool=query_chem_tool, tool_input={'plant': '博落木'}

Observation: {'result': '博落木的化学成分包括挥发油和酚类化合物。挥发油主要成分为柠檬烯，酚类化合物主要有没食子酸、香豆素、原儿茶酸和原儿茶酚等。'}

Final Answer: 治疗皮炎的植物包括流苏子和博落木。流苏子的主要化学成分包括抗皮炎、抗氧化和抗炎等药理活性成分。其用途是根据其根辛辣，可治疗皮炎。博落木主要用于治疗皮肤疾病，如慢性眼皮炎和去除疣。其化学成分包括挥发油和苯类化合物，主要成分有槲皮素、香豆素、原儿茶酸和原儿茶酸酯等。

用户可**在线上传PDF格式**的研究论文，系统运用基于大语言模型的**文献知识智能抽取**技术，智能化地识别并提炼出论文中的核心内容。同时，用户还可以通过人工智能对话的方式进行**深度追问**，针对论文细节进行探讨，获取更深入的理解和洞察。论文研读功能助力用户迅速把握论文的精髓，大幅提高文献阅读的效率。

The screenshot displays a research paper titled "Approaching a stable oxygen redox reaction in lithium-rich cathode materials: structural perspectives from mechanism to optimization" from the Journal of Materials Chemistry A. The paper is a review by Zijia Yin, He Zhu, Yalan Huang, Dong Luo, Yang Ren, Si Lan, and Qi Liu. The interface includes a sidebar with navigation options and a main content area with the paper's abstract and introduction. On the right, a chat interface provides AI-generated insights and summaries.

你可能想要了解这篇文章的
“研究主题”，“研究目标”，“研究方法”，“研究技术”，“创新点”，“研究结论”，“研究改进方向”

User: 文章的“研究方法”

研究目标概括如下:

1. 分析LiTMO₂结构中四面体位点的锂离子迁移，对高电压下的相变缓解作用。
2. 探究LLO材料的电化学充放电机制，特别是4.5V以上区域的氧化还原过程。
3. 研究LLO材料的氧释放与TM迁移、晶格收缩、相变的关系。
4. 探讨LLO材料的结构改性策略，如调节TM-O键、构建异质界面等，以提高氧还原反应的可逆性。
5. 比较O3型和O2型LLO材料中TM离子的可逆迁移行为。
6. 通过Li-Na离子交换策略制备O2相LLO材料，实现TM离子的可逆迁移。
7. 总结LLO材料的结构改性策略，讨论其优势与局限性。

User: 文章的“创新点”

1. 锂富集层状氧化物(LLO)材料的阴离子氧化还原反应可提高比容量，但不可逆反应导致结构退化。
2. 阴离子氧化还原过程涉及氧空位形成、过渡金属离子迁移和结构相变，导致容量和电压衰减。
3. 改善策略包括调控TM-O键、异质界面设计、表面缺陷工程和O2相构建，以抑制不可逆反应。
4. O2相结构可促进过渡金属离子的可逆迁移，抑制不可逆反应，稳定电压和容量。
5. 各种改性策略各有优劣，需进一步优化以实现商业化应用。

User: 请用英文来回答

Here are some key takeaways from the literature on lithium-rich layered oxide cathodes:

- **Anionic Redox Reaction**: The unique anionic redox reaction in the bulk phase contributes to the high capacity of lithium-rich materials, while the irreversible anionic redox reaction on the surface leads to structural rearrangement and irreversible capacity loss.
- **Structural Evolution**: Charging above 4.5 V leads to oxygen release, TM migration, and phase transitions, resulting in irreversible capacity fade and voltage decay.
- **Modification Strategies**: Strategies like regulating TM-O bonds, surface coatings, and constructing O2 phases aim to stabilize the structure and enhance reversibility, with O2 phases showing promise for enabling reversible TM migration.
- **Challenges**: The dynamic coupling between OR reaction, TM migration, and electronic structure during cycling is not fully understood. Efficient synthesis methods for modification strategies are also lacking.
- **Future Research Directions**: Further research is needed on oxygen loss mechanisms, TM migration paths, and the development of efficient modification strategies using advanced characterization techniques and theory calculations.

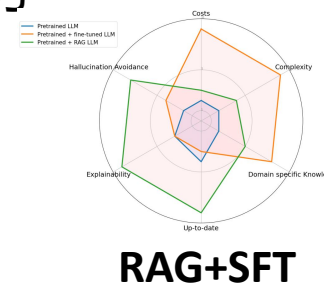
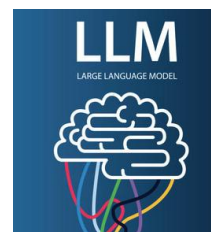
Type your message here Submit Re-upload pdf file

AI大模型与科学大数据、关联数据的融合应用，通过语义知识库构建、大模型检索增强与训练等关键方法，全面提升科学数据智能化服务的准确性和可解释性，协助各领域的科研人员以更高效的方式进行数据资源深度检索、挖掘与知识发现服务，助力数据中心资源价值倍增。

RAG与微调技术优势互补，RAG集成外部数据，提升回答的准确性和可解释性，微调提升模型完成特定任务的能力与性能。未来两者相互配合，共同为科学数据赋能，提高研究深度与广度，推动科研创新。

未来应用方向：

- 多模态数据智能分割
- 知识表征协同优化
- 检索重排序
- 上下文窗口动态扩展
- 场景自适应提示优化
- 问题语义重塑增强
- 高效并行推理加速
- 通过微调增强模型能力





中国科学院计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

感谢聆听!

