教师简介

窦志成 长聘教授

博士生导师、副院长，中国计算机学会大数据专家委员会秘书长，中国中文信息学会理事、信息检索专委会副主任，TOIS 期刊副主编、中国数据大会程序主席，中国大数据学术会议组织主席，全国信息检索学术会议大会主席。主要研究方向为信息检索、大模型、智能体、大模型检索增强、AI 搜索、司法智能等。在国际知名学术会议和期刊上发表论文 200 余篇，带领团队研发涉外法治大模型，开源大模型检索增强工具包 FlashRAG、信息智能体系列 (WebThinker、ARPO、DeepAgent 等) 累计获得 GitHub 星标 8800 余枚 (截至 2025 年 11 月）。曾获教育部自然科学奖一等奖、国际信息检索大会最佳论文提名奖、国际万维网大会亮点论文奖、亚洲信息检索大会最佳论文奖、全国信息检索学术会议最佳论文奖等奖励。与华为、字节、OPPO、腾讯、快手、百度、百川、智源、微软亚洲研究院等单位开展科研合作。

研究方向

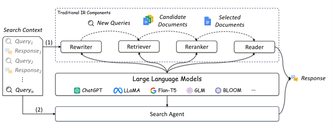
信息获取

01	检索与推荐模型	排序模型、个性化搜索、多样化搜索、对话式搜索、商品推荐、新闻推荐
02	大模型检索融合	大模型赋能检索、生成式检索、大模型检索增强、嵌入模型、AI搜索
03	信息智能体	智能体、深度搜索、深度研究、多智能体、强化学习、工具调用、记忆机制
04	个性化信息获取	个性化搜索、个性化对话、个性化推荐、个性化LLM、用户建模、个性化记忆
05	多模态信息获取	多模态表征模型、多模态与跨模态检索、多模态检索增强、多模态智能体
06	司法智能与AI治理	司法智能体、判决预测、涉外法治大模型、生成内容识别、大模型价值观对齐

研究成果　新一代智能信息检索

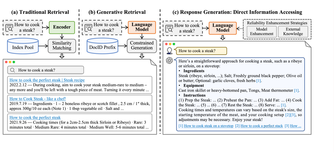
1. 大模型赋能信息检索：从组件优化到范式变革

大模型驱动信息检索范式变革



✔大模型优化现有搜索引擎中的查询理解、检索、排序、答案生成等模块，进一步提升搜索质量

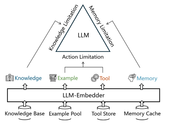
大模型赋能信息检索综述：系统性梳理 350+ 篇文献，谷歌学术引用 620+ 次
Large Language Models for Information Retrieval: A Survey, ACM TOIS



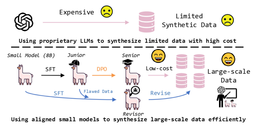
✔传统的匹配检索范式过渡到以大模型为基础架构的生成式检索范式；
✔传统的文档检索范式过渡到端到端的答案生成范式

生成式信息检索综述：系统性梳理 370+ 篇文献，谷歌学术引用 150+ 次
From Matching to Generation: A Survey on Generative Information Retrieval, ACM TOIS

通用嵌入模型 (Embedding)

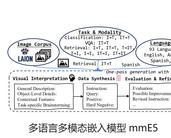


LLM-Embedder: 面向 LLM 的嵌入模型

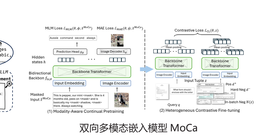


SPEED: 小模型合成高质量嵌入数据

文本嵌入模型 LLM-Embedder/SPEED: Hugging Face 共下载 200 余万次
A Multi-Task Embedder For Retrieval Augmented LLMs, ACL
Little Giants: Synthesizing High-Quality Embedding Data at Scale, NAACL



多语言多模态嵌入模型 mmE5

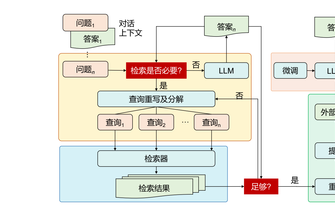


双向多模态嵌入模型 MoCa

多模态嵌入模型 mmE5/MoCa: Hugging Face 共下载 3 万余次
mmE5: Improving Multimodal Multilingual Embeddings via High-quality Synthetic Data, ACL Findings
MoCa: Modality-aware Continual Pre-training Makes Better Bidirectional Multimodal Embeddings, arXiv

2. 大模型检索增强（RAG）：打通大模型落地的最后一公里

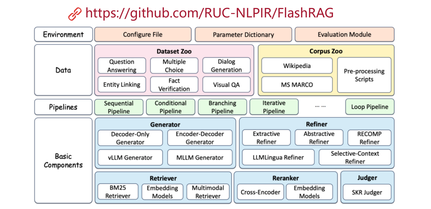
大模型检索增强系统化研究



围绕 RAG 架构以及对话上下文理解、意图识别、检索器、重排器、提炼器、模型微调等 RAG 组件开展深入研究

开源大模型检索增强工具包 FlashRAG

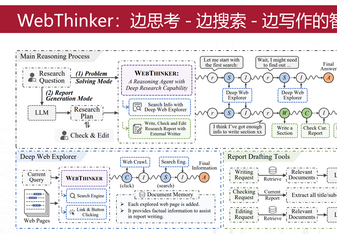
https://github.com/RUC-NLPIR/FlashRAG



集成 18 种 RAG 方法，7 种信息智能体方法；40 个数据集，累计下载 8.4 万余次
获开源社区 GitHub 星标 3.2K+，单日趋势榜第三

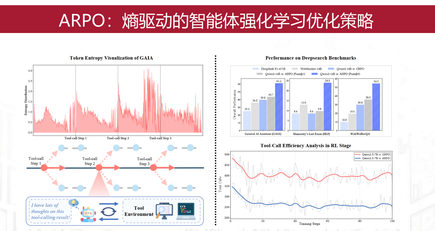
3. 深度搜索及研究智能体　https://github.com/RUC-NLPIR/iAgent

WebThinker：边思考 - 边搜索 - 边写作的智能体框架



获开源社区 GitHub 星标 1.4K+，Hugging Face 论文日榜第一
WebThinker: Empowering Large Reasoning Models with Deep Research Capability, NeurIPS

ARPO：熵驱动的智能体强化学习优化策略



获开源社区 GitHub 星标 820+，Hugging Face 论文周榜第一
Agentic Reinforced Policy Optimization, arXiv

联系方式

电子邮箱：dou@ruc.edu.cn
个人网址：http://playbigdata.ruc.edu.cn/dou/
Lab 网站：https://ruc-nlpir.github.io/

个人网站　　学院主页