

实践活动三

基于检索增强生成（RAG）的知识检索与信息整合实践

本实践活动围绕检索增强生成（Retrieval-Augmented Generation, RAG）技术在科研与信息处理场景中的应用展开，重点探索如何将外部文档与大语言模型生成能力相结合，以提升信息获取的可靠性与可追溯性。在前期实践中发现，单纯依赖大语言模型进行内容生成，容易受到模型训练语料与生成策略的影响，难以保证输出内容始终基于真实、可验证的信息来源。

基于上述问题，本实践引入了 RAG 技术作为知识获取与生成过程中的支撑机制。在实践过程中，首先对研究资料和文档数据进行了系统化处理，包括文档收集、文本清洗以及内容切分等步骤。通过将长文本拆分为若干语义相对完整的文本单元，为后续向量化与检索提供基础。

在检索策略方面，本实践探索了向量检索与关键词检索相结合的方式，以提高检索结果的覆盖度与相关性。在实际应用中发现，单一检索方式在面对复杂研究问题时容易出现偏差，而混合检索能够在一定程度上弥补不同方法的不足。检索得到的结果并未直接作为最终答案，而是作为“证据上下文”输入至大语言模型，用于辅助生成内容。

在实践过程中，重点关注了检索结果与生成内容之间的关联程度。通过多轮测试发现，若不对检索结果进行筛选与结构化组织，模型容易忽略关键信息或误用无关内容。因此，在系统设计中，对检索结果进行了整理与过滤，使其以更加明确的上下文形式参与生成过程。

通过本次实践，逐步认识到 RAG 技术在科研场景中的核心价值并非“自动给出结论”，而在于为模型生成提供可靠的信息基础。该实践加深了对大模型与外部知识协同机制的理解，也为后续更复杂的研究辅助系统设计提供了重要经验。