arxiv.org/abs/2507.14189

DeepWriter: A Fact-Grounded Multimodal Writing Assistant Based On Offline Knowledge Base

Song Mao[1], Lejun Cheng[1,2], Pinlong Cai[1], Guohang Yan[1], Ding Wang[1], Botian Shi[1],

未开源

[1]Shanghai AI Lab, [2]Peking University

Correspondence: caipinlong@pjlab.org.cn

简介

本文提出DeepWriter，一个面向金融、法律等专业/垂直领域的报告生成系统，特点是基于本地文件(不联网搜索)、具备多模态写作能力并且严格依据资料事实。 简单来说，DeepWriter 包含本地文件处理和在线报告生成两个模块：在离线阶段，系统对非结构化文档(主要是pdf文件)进行解析与分块(chunk)，提取文本、表格、图像等内容，并构建三层级知识库(文档-页面-chunk)进行知识管理和存储；在线阶段，系统接收query后，依次执行任务拆解、大纲生成、分段写作、图文融合与细粒度引用，最终生成一篇结构清晰、语义连贯、信息可溯的专业报告。总的来说，DeepWriter适合对内容准确性要求高、图文信息丰富的长文写作任务。

背景

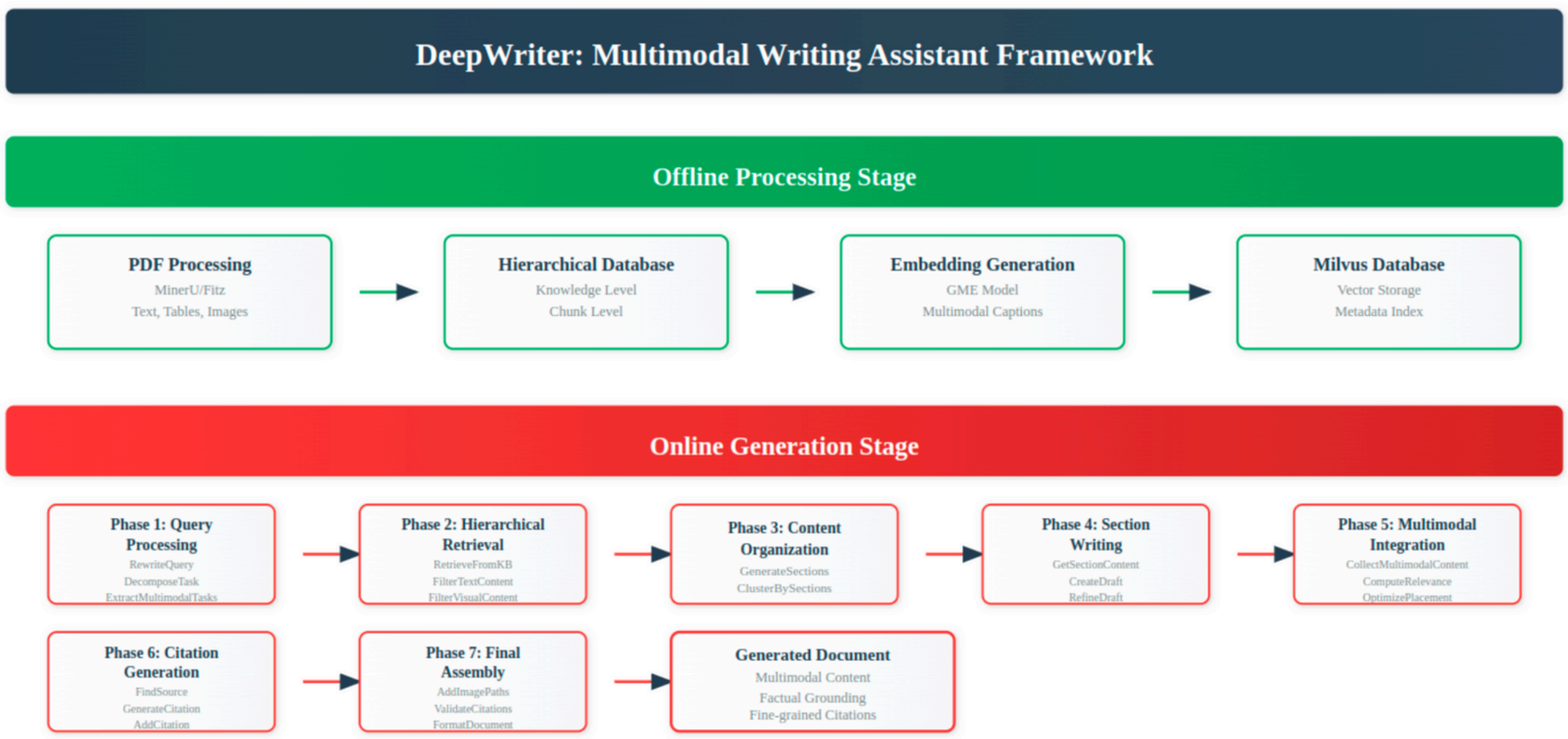
对于金融、法律等专业领域的报告生成任务，目前基于llm的主流方案主要有两类：1) RAG ，虽然可以引入外部知识，但在生成长篇报告时常出现上下文不一致的问题，且难以实现精准的资料引用(citation)，尤其在处理图表等多模态内容时效果也不好；2) llm+search engine，但由于网络内容质量参差不齐，最终生成结果往往缺乏可信度与专业性。为此，本文提出一种新的解决方案：基于本地数据集构建一个具备多模态写作能力并且严格依据资料事实的垂直领域写作助手。

实验设置

- 使用MinerU处理pdf文件
- 向量化模型：gme-Qwen2-VL-2B-Instruct
- 向量数据库：Milvus
- 生成图和表格caption模型：Qwen2.5-VL 7B
- 执行写作任务的模型：Qwen2-7B
- 验证集：WTR (World Trade Report)
- llm-as-a-judge模型：Prometheus2-7B

DeepWriter架构

- 本地文件处理：主要是pdf文件，进行结构化解析，提取出文本、表格、图像等多模态内容，并结合元数据构建三层级知识表示
- 在线生成阶段：根据query，首先对query重写与任务拆解，明确生成目标，然后通过大纲规划、内容聚类与分段写作、图文内容的匹配与位置优化、细粒度引用



思考

论文读下来，感觉DeepWriter的内容生成策略是偏保守的，但是这好像恰恰是它的优点，因为作者考虑的应用场景就是针对垂直领域文件的知识管理、知识重组与报告生成，保守设计是为了最大程度规避幻觉问题，特别是多模态插图和段落级引用机制是很大的亮点，我在用ChatGPT时就会不自觉的信任那些给出citation的内容。当然了，由于没开源，具体使用的效果如何就不清楚了，总体上本文提出的针对垂直领域的专业生成思路，还是挺好的。