

Multimodal DeepResearcher: Generating Text-Chart Interleaved Reports From Scratch with Agentic Framework

如何生成包含可视化图表的技术报告? arxiv.org/abs/2506.02454

Zhaorui Yang^{§*}, Bo Pan^{§*}, Han Wang^{§*}, Yiyao Wang[§], Xingyu Liu[§]
Minfeng Zhu^{‡☒}, Bo Zhang^{§☒}, Wei Chen[§]

[§]State Key Lab of CAD&CG, Zhejiang University [‡]Zhejiang University

简介

注意：本文的多模态是文本和图表，不是文本和图像

为解决DeepResearch 报告(report)中缺乏可视化图表(chart)的问题，本文提出了Multimodal DeepResearcher，简单来说，它是prompt engineering驱动的Agent框架，workflow分为4个阶段：Researching (根据query利用search engine检索相关素材并归纳)、Exemplar Report Textualization (对于prompt中few-shot示例报告中的chart，用语言描述chart内容)、Planning (生成报告大纲和chart风格指南)、Multimodal Report Generation (报告生成)。

本文一个重要的点在于few-shot和Formal Description of Visualization (FDV)，作者将人撰写的包含图表的报告以few-shot的方式放入prompt，期望让llm学习报告的结构、风格和如何插入图表。为了让llm能准确理解和生成图表，提出了FDV，就是用语言对图表做多方面描述，这样报告就转为纯文本了，便于llm理解报告内容。同时模型在生成报告时对于图表先生成FDV，再让llm根据FDV生成D3.js代码，最终得到包含文字和可视化图表的报告。

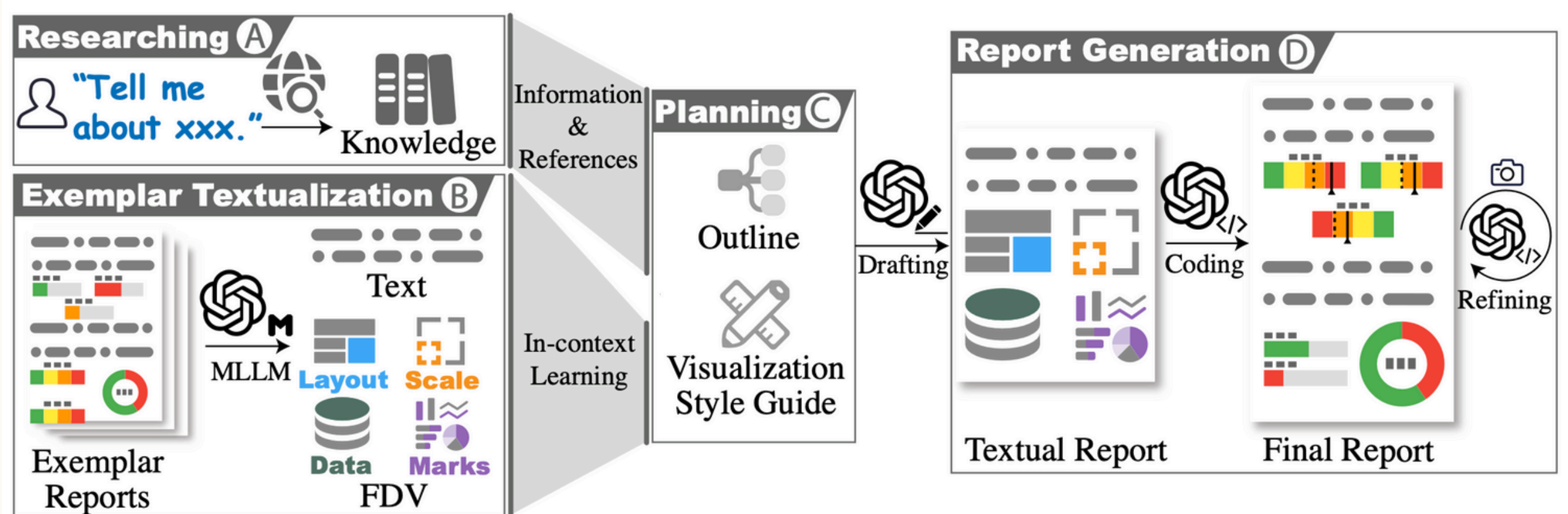
背景

现有的Deep Research研究的生成报告(report)基本上都是纯文本内容，而人写的报告往往会在数据描述部分插入图表(chart)，毕竟一图胜千言，因此本文考虑如何在报告中加入可视化图表(chart，比如柱状图、折线图、散点图等)，这样生成的报告更具有可读性。

实验设置

- search tool: Firecrawl API
- 推理和planning 模型: GPT-4o-mini
- 生成FDV的模型: Claude 3.7 Sonnet
- 生成报告模型: Claude 3.7 Sonnet

Multimodal DeepResearcher



思考

目前长文生成领域，学术界有不少工作聚焦在故事生成，从DeepResearch角度，我更关注如何生成技术报告，而专业性的技术报告缺少不了图表📊，本文的研究角度很有应用价值。希望尽快开源代码吧:)