

作业 2

基于结构化数据的报告/报表自动生成系统设计与实现

一、作业背景

在现代医疗、护理、行政等管理场景中，大量信息需要被定期整合为可读性强、结构清晰、带图表的报告或报表。如何从多源数据中自动提取关键要素，支持用户定制化生成包含图表、汇总、引用、鱼骨图、流程图等内容的总结报告，是医院信息化系统建设的重要组成。本作业旨在设计一套通用的 AI 报告/报表自动生成系统，实现结构化数据到可视化网页报告的闭环转化。。

二、数据说明

1. 输入与输出

输入：

可以参考 `input_data.zip` 里的知识文件、表格文件等作为原始输入数据，也可以在各种网站或者资料作为用于生成报告的原始数据。（自行扩展的数据需要提交并在报告中说明）

输出：

输出包含中间文件 `json`（中间文件生成格式参考 `mid_output_json.js`）以及前端展示文件

2. 基本实现示例

`simple_exmaple_IO.zip` 为本作业的基本实现示例（实现类似结果可得基础分）

效果：用于实现“生成 2025 年 5 月份护理部的理论知识培训”报告

输入：`/simple_exmaple_IO/2025 年 5 月护理部理论知识培训.docx`

输出（中间文件）：`/simple_exmaple_IO/mid_output.json` 为使用智能报告系统得到的中间解析输出

输出（结果展示）：`/simple_exmaple_IO/final_output.html` 为最终展示报告

三、任务要求

1. 支持用户通过自然语言或模块化方式输入报告生成请求（如：“生成 5 月份护理质量报告”）
2. 构建统一的中间 JSON 结构格式（可覆盖不同类型报告模块）
3. 支持以下模块的可视化报告输出：
 - * 表格（数据总览、评分情况等）
 - * 图表（饼图、柱状图、折线图等，展示统计趋势）
 - * 鱼骨图（根因分析图）
 - * 流程图（整改流程、操作步骤等）
 - * 文字总结/引用段（自动提炼问题、分析建议）
4. 支持输出 HTML 格式的前端报告页面
5. 提供至少 2 类不同类型的场景输入，生成两份报告示例

四、提交内容

需提交以下内容：

1. 项目技术报告 PDF，包含：
 - 报表结构抽象设计（模块划分与中间结构定义）
 - 中间 JSON 格式示例与生成逻辑
 - HTML 报表模板设计说明（图表渲染方式、布局逻辑）
 - 示例输入与输出展示（JSON 文件 + HTML 页面截图）
 - 系统自动图形模块（鱼骨图、流程图）的绘图方式说明
 - 系统优势、适配场景、扩展方向说明
2. 代码包：
 - 中间数据生成模块
 - HTML 报告模板代码
 - 自动绘图模块（流程图、鱼骨图生成器）
 - 示例数据与运行说明（README.md）
3. 示例结果：
 - 至少兼容 2 种类型输入（如 pdf、excel 数据作为输入）分别得到的完整中间 JSON 和 HTML 输出

五、评分标准

1. 代码部分（70 分）

A. 功能能力实现（50 分）

项目	说明	分值
输入解析与数据识别能力	能自动从用户输入中识别关键信息，生成对应中间结构	10
中间结构设计合理性	报表结构抽象能力强，具有通用性与可扩展性	15
图表渲染与数据准确性	基础图表样式清晰，数据正确映射，结构合理	15
特殊图绘制能力	能正确生成流程图/鱼骨图，结构合理	10

B. 代码结构与工程规范（20 分）

项目	说明	分值
模块划分合理性	项目结构清晰、职责明确、模块之间解耦	10
代码规范性与可读性	命名清晰、注释充分、逻辑简洁，便于理解与复现	10

2. 报告（30 分 + 10 分 bonus）

项目	说明	分值
报告内容完整性	包含设计方法、中间格式说明、示例输出等	15
报告表达清晰度	结构条理清晰，语言准确规范，有条理地表达设计思路与技术路线，内容充实	15
附加功能或创新分析	支持模块自定义、自动摘要生成、页面自适应，支持打印或导出（如 PDF）	10

六、参考建议

推荐使用的工具或技术（也可使用其他方法）包括：

文本解析与结构抽取技术

1. 使用 python-docx、python-pptx、PyMuPDF、pdfplumber 等工具从 Word、PPT、PDF 中提取原始内容（文本、表格、标题等）。
2. 对 Excel 或 CSV 表格数据，使用 pandas、openpyxl 进行清洗与结构转换。
3. 结合 jieba、spaCy、HanLP 或 transformers 模型完成关键词提取、实体识别、摘要提炼等。

模板构建与内容生成

1. 使用 Jinja2、Handlebars.js 等模板引擎将中间 JSON 数据填入 HTML 模板，实现内容动态渲染。
2. 前端页面可用 Vue、React 配合 TailwindCSS 或 Bootstrap 实现响应式布局与样式美化。
3. 推荐用配置文件（如 JSON Schema 或 YAML）定义报告模块结构、显示顺序和内容模板。

图表绘制与可视化

1. 使用 Chart.js 或 ECharts 构建统计图，如柱状图、饼图、折线图等，展示分析数据。
2. 对于个性化图形需求，可用 D3.js 或 SVG/canvas 构建自定义图表（如多层分布、时间线等）。
3. 通过 html2canvas 或 jsPDF 实现 HTML 页面导出为图片或 PDF 报表，支持打印。

流程图与鱼骨图绘制

1. 使用 Mermaid、Flowchart.js 或 PlantUML 自动渲染整改流程图。
2. 鱼骨图可通过 HTML+CSS+SVG 结构手动绘制，也可将 JSON 映射为树结构并渲染为 canvas 图形。
3. 更复杂的图形展示可考虑引入 GoJS 或 X6 等图形引擎，支持交互式节点图。

AI 辅助模块与智能扩展

1. 利用 ChatGPT、DeepSeek、ERNIE 等模型从原始文本中自动提炼问题分析段、改进建议。
2. 结合提示词工程或模板填空技术，构造“总结性段落”、“引用性归纳”、“风险建议”等内容块。
3. 借助 LangChain、Semantic Kernel 等框架，实现用户用自然语言下达“生成报告”指令。

4. 使用语义搜索或问答模块（如 Haystack、LlamaIndex）支持对报告内容的智能查询。

可结合课堂内容、已有代码库或开源工具，自主选择实现路径。