

作业

基于多模态课件内容的个性化试题自动生成系统设计与实现

一、作业背景

随着人工智能在教育场景中的深入应用，如何对教学资源进行自动理解、提炼知识，并基于个体差异生成针对性评估内容，成为智能教育的重要研究课题。本作业旨在构建一个自动化系统，实现从教学课件中抽取知识并生成个性化测试题，融合自然语言处理、多模态内容理解与教育任务建模能力。

二、数据说明

参考数据 data.zip，解压后结构如下：

- /data/video: 护理类教学视频
- /data/ppt: PPT 培训课件
- /data/pdf: 护理方案与操作文档
- /data/word: 文本知识文档

上述资料可作为知识抽取与试题生成的输入来源（也可自行找寻其他类似的课件/教材，自行扩展的数据需要提交并在报告中说明）。

三、任务要求

设计并实现一个原型系统，完成以下目标：

- 对教学视频、PPT、PDF 等课件内容进行解析与文本提取
- 抽取关键知识点或主题信息
- 自动生成多种题型的测试题，包括选择题、填空题、问答题等
- 提供前端页面展示（非强制要求，**加分项**）

四、提交内容

需提交以下内容：

- 项目技术报告（PDF），内容应包含以下部分：
 - 数据预处理流程与工具选择
 - 核心模块功能设计与技术路线（含文本解析、知识点抽取、试题生成方法）
 - 示例输出结果（包括至少 15 道题、2 种以上题型），并标注来源资料
 - 实验测试分析与存在问题
 - 模块接口设计与使用说明（如有 API 或前端）
- 完整代码包：
 - 包括全部源代码、依赖库说明、运行方式说明（README）
 - 可选提供运行脚本或 Jupyter Notebook 演示入口
- 示例输出题目文档（结构清晰、可阅读、附答案）

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
内容提取能力	能从视频、PPT、PDF 中准确提取出文字信息	15
知识组织合理性	能将提取内容进行简单分类或结构化表示	15
题目生成类型	能自动生成多种题型，题目表达清晰	10
题目数量及格式	提供不少于 15 道题目，格式整齐	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	包含方法设计、数据处理、关键模块说明、结果展示（生成的试题）、遇到的问题及分析	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	对非强制功能（如前端、算法优化、跨模态协同）进行设计说明或实验分析	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：

- 文本提取工具：python-docx, python-pptx, PyMuPDF, pdfminer.six
- 视频转文本技术：Whisper, AliASR 等开源语音识别模型
- 语言模型接口：OpenAI GPT, DeepSeek, 通义千问 等
- 试题生成方法：命名实体识别（NER）、关键词提取（TextRank）、模板填空、提示词工程等

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。