PYTHON 测试题 1

## 技术面试题

PYTHON 测试题	1
JAVA 测试题	
前端技术面试题	
白然语言和人工智能测试题	Δ

## PYTHON 测试题

#### 需求:

1,对 WORD 和 PDF 进行文档分析,类似下面的效果,可以导出 markdown 文件



### 技术选型:

- Word 解析: python-docx (读取文本、表格、图片)
- PDF 解析: PyMuPDF (文本、图片提取)、pdfplumber (表格提取)
- Markdown 生成: 手动拼接或 markdown 库
- 图片处理: PIL(保存图片文件)
- 其他工具: os (路径管理)、re (正则清洗文本)

JAVA 测试题

## JAVA 测试题

#### 需求

1、B/S 结构, 部署在自己申请的云服务器上(免费和极低价格的)

- 2、数据库用 MYSQL
- 3、建一个数据管理模块,可以上传 EXCEL 文件, 创建各个数据表
- 4、 建一个 CORELDRAWX6 的模板管理模块,管理 CORELDRAW 的模板,每个模板里面的变量和数据表里的字段建立关联

开发CorelDRAW X6插件需要掌握以下知识和技能:

- 1. VBA编程: CorelDRAW X6插件开发可以利用Visual Basic for Applications (VBA)进行编程。需要了解VBA的基本语法和编程结构,以及如何在CorelDRAW中使用VBA进行自动化操作 ??。
- 2. CoreIDRAW对象模型: 了解CoreIDRAW的对象模型是开发插件的基础,包括文档 (Document)、图层 (Layer)、页面 (Page)、形状 (Shape) 等对象的操作和属性 22。
- 3. CoreIDRAW API: 熟悉CoreIDRAW提供的API, 这些API允许开发者访问和控制CoreIDRAW的内部功能,如创建形状、填充颜色、操作图层等 "。
- 4. 插件架构: 了解如何创建和管理插件,包括插件的打包、部署和与CorelDRAW的交互方式。特别是GMS(Corel Macro System)文件的创建和管理 "。
- 5. 用户界面设计:插件可能需要提供用户界面,因此需要了解如何在CorelDRAW中创建和自定义用户界面元素,如工具栏、按钮和对话框 "。
- 6. 调试和测试:掌握如何在开发过程中调试和测试插件,确保插件的稳定性和兼容性 "。
- 7. C#编程: 对于更高级的插件开发,可能需要使用C#语言,特别是涉及到CorelDRAW CPG (Corel Plug-in Gallery) 扩展开发时 ?? 。
- 8. SDK和开发工具:需要获取CorelDRAW SDK(软件开发工具包),这是开发插件所必需的,包括库文件、示例代码和文档 "。
- CQL语法: 了解CoreIDRAW查询语言(CQL),这是一种用于查询和操作CoreIDRAW中对象的语法
- 10. 编译和部署: 掌握如何编译插件代码,并将其部署到CorelDRAW中供用户使用 🤊 。
- 5、选择一个工作表和 CORELDRAW 的模板,实现自动把数据传到指定 CORELDRAW 的模板并生成一个 CORELDRAW,然后可以下载

前端技术面试题

# 前端技术面试题

根据以下资源,开发一个 AR 的应用

https://github.com/AR-js-Org/AR.jS

https://github.com/aframevr/aframe

部署在自己本地机后联系我通过远程视频会议看下效果

## 自然语言和人工智能测试题

需求: 比较两个 PDF 的差异 (PDF 可能是图片式)

- 1、 上传两个要比较的 PDF
- 2、 点比较后进行比较
- 3、 呈现差异结果显示
  - 1) 可视化差异展示
  - 2) 生成差异报告:将比较结果生成详细的差异报告,包括文本内容的增 删改、格式变化等,方便用户查看和理解。
  - 3) 可视化界面: 开发可视化界面,以直观的方式展示两个 PDF 的差异,如高亮显示不同之处、提供侧边栏对比等。

#### 参考:

#### 一、文本内容比较

- 1. 文本抽取与对齐
  - o OCR 技术: 处理扫描版 PDF,将图像文本转换为可编辑的文本。
  - o **文本分句与对齐**:通过 NLP 分句技术将两篇文档的文本对齐,逐句或逐段 比较差异。

### 二、文档结构分析

- 1. 布局与格式比较
  - o PDF 解析库(如 PyMuPDF、PDFMiner): 提取文本、表格、图像的位置信息。
  - o **计算机视觉:** 识别页面布局(如标题层级、段落间距)和格式变化(如字体、 颜色)。
- 2. 表格与图表处理
  - o 表格识别:用 CV 模型(如 TableNet)提取表格数据并比较内容。
  - o **图表差异检测**:通过图像对比技术(如哈希算法、特征点匹配)检测图表的 修改。

### 三、高级 AI 技术

- 1. 差异摘要生成
  - o 使用生成式 AI 总结文档差异,输出结构化报告。
- 2. 异常检测
  - o 训练分类模型识别敏感内容的增删(如合同条款、法律声明)。
- 3. 多模态融合

o 结合文本、图像、表格等多模态数据,全面分析差异。

## 四、工具与实现建议

### 开源工具

- o **Python 库:** diff-match-patch (文本差异)、PyPDF2 (PDF 解析)、Spacy (NLP 分析)。
- o AI 框架: Hugging Face Transformers (预训练模型)、YOLO (图像检测)。

测试样本文件: PDF 差异编辑测试样本. rar

链接: https://pan.baidu.com/s/1h6RPvfXzpvzjB4C7Jp0p7w 提取码: p3e3