**课程项目开题报告**

题目：智能简历分析系统

成员：

2021302111169 牛锴鹏

2021302111252 毛子昊

2021302111380 杨力闻

2021302111275 林诰人

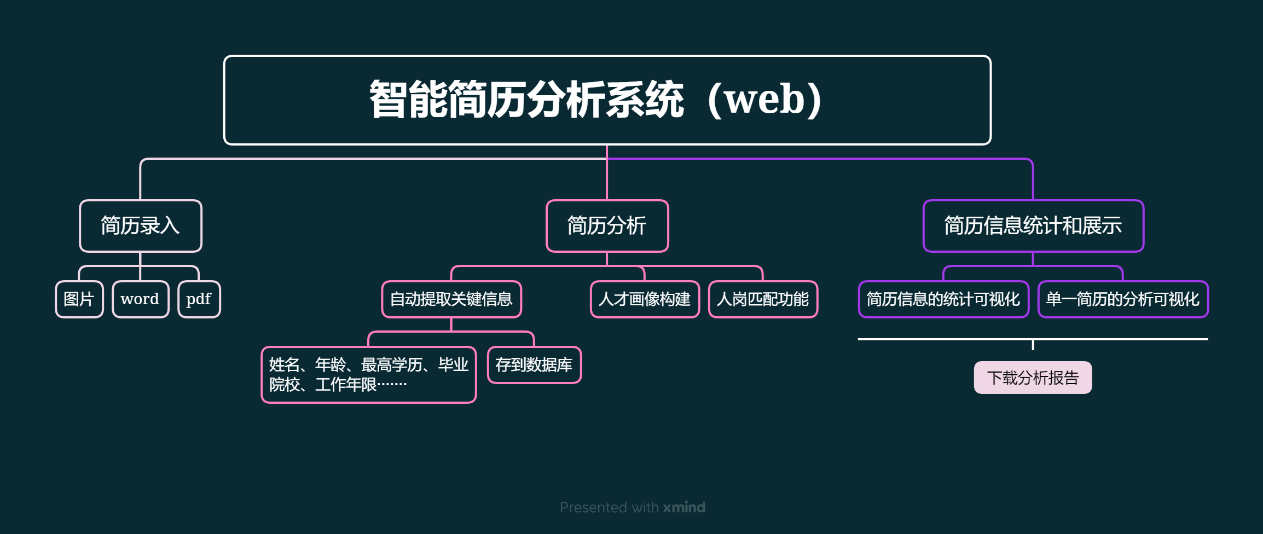
## 一、选题动机

在当前激烈的招聘市场中，企业和求职者都面临着巨大挑战。传统的人力招聘方式不仅效率低下，耗时较长，而且容易出现人为判断失误。因此，开发一款能够提高招聘效率、降低人力成本的智能简历分析系统具有重要的实际意义。该系统可以帮助企业更高效地筛选合适的人选，提升招聘质量，同时为求职者提供更好的匹配机会。本项目将涉及自然语言处理、文本分析、图片文字识别等多个技术领域，具有一定的技术挑战性。

随着人工智能技术的普及，市场对自动化、智能化招聘解决方案的需求越来越大。通过实现简历信息和岗位需求的自动分析，智能简历分析系统将在提高招聘效率和质量方面发挥关键作用。此外，简历分析结果的可视化展示将帮助企业更直观地了解分析结果，从而提高企业体验。本系统还将支持多种简历格式输入，包括文本、PDF、DOCX以及图片格式，为用户提供更为便捷的简历录入方式。

（我们计划拿这个项目去参加中国软件杯，所以选题动机有参考中国软件杯的题目需求）

## 二、软件功能



## 三、技术路线和难点

### 前端

**技术路线:**

前端使用Vue全家桶 + Nodejs + ElementUI 开发，具体功能如下：

1. ***登陆页面:***

用户提交用户名和密码，前端校验用户名和密码有效性后，调用后端接口/login。后端处理请求，向前端返回用户token。

前端将接收到token保存在浏览器cookie中，以保存用户登录状态。首先判断用户是否登录获得token。若无token且访问页面不在免登录白名单中，则重定向至登录页；若有token，则判断用户是否已经获取用户信息（如用户权限等信息），若未获取用户信息，则使用登录获得的token调用后端接口获取用户名、权限等用户信息，并保存在vuex全局的store中，便于统一管理。获取完用户信息之后，基于用户权限和页面权限生成用户能访问的路由表, 然后通过vue-router动态添加可访问路由表，最后根据用户权限显示可使用的功能和页面。

1. ***注册页面:***

用户进入注册页面时，提交用户名、密码、所属企业和个人简介，前端校验输入有效性后，调用后端接口进行注册，等待数据库管理员审核.

1. ***登出***

用户点击退出按钮，前端调用后端接口, 清除浏览器的token，重置路由表和Vuex全局的store，最后跳转至登录界面，完成登出。

1. ***Token***

用户登录系统后，每次向后端接口发送请求时，若用户token存在，前端的请求拦截器都需要在请求头的Authorization字段加上用户token，便于发送请求时携带。后端校验用户token的有效性，判断其是否登录或登录状态是否过期。

1. ***简历信息展示***

获得后端传递的json数据（包含图表数据json、分析报告json、岗位报告json等），分类进行可视化处理。

**技术难点：**

1. ***响应式：***

参考了类似功能的网站，发现响应式设计效果不好，考虑可能原因是简历信息的复杂性导致的，因此设计页面的响应式是一个难点问题。

1. ***多端适配：***

设想使用场景为PC端+移动端（因为题面要求WEB / App形式），因此很可能需要进行多端适配来适应移动端使用。

1. ***部署与CDN：***

由于简历数据量天然较多，且页面较为复杂，为了提高渲染速度，应使用CDN进行内容分发，因此需要服务器资源和注册域名备案。

### 后端

**技术路线**

使用C#编写web后端，要考虑的技术难点有异步编程，数据库访问，接口设计，身份验证和授权，性能优化，安全性。

**技术难点**

①首先对于高并发，多线程：对于一个web应用程序，就一定要考虑程序的并发量，吞吐量，响应速度等。所以，此时，可以使用多线程来提高效率，但我们打算使用异步编程来解决并发的问题，但是异步编程是一个比较复杂的技术，是一个难点。

②数据库设计：说到web程序开发，自然离不开数据库的应用。一个好的数据库的设计对于web程序的性能的提高是十分明显的。数据库的设计，首先就是得考虑数据库的存储结构，查找性能，数据库的完整性约束，数据库安全等。所以数据库的设计是一个难点。

③接口设计：web后端首先得响应web发来的请求，以及调用后台AI算法来演算出正确的简历分析，因此好的接口设计对一个web后端的开发是比较重要的，他得即考虑前端请求，又得调用AI分析算法，是一个重难点。

④身份验证和授权：对于一个web后端来说，应当有一个好的身份验证和授权功能，以及用户的功能权限有所区分。

⑤安全性：作为一个简历分析系统的web后端，安全性是十分重要的，因为简历设计个人隐私和招聘公司的利益，所以此时安全性问题是十分重要的。但同时安全性问题又是一个难点，得需要我们掌握很多关于安全技术。首先，我们的web应用程序得防止各种安全攻击，例如XSS，CSRF，SQL等，需要我们对很多相关技术的了解（如输入验证，输出编码，防御攻击等）。

总而言之，上述是我们小组对后端设计的技术路线和重难点的思考，但是肯定是不太完善的，问题永远肯定会在实际编程的过程中出现，但我们小组觉得，我们可以解决这些重难点。

### 算法

算法方面主要涉及两部分，一部分是分析简历，提取信息；另一部分是人岗匹配。

分析简历方面：

技术路线：1. 对于文本格式的简历，利用自然语言处理（NLP）技术进行关键信息提取和人才画像构建。使用现有的NLP框架（如spaCy、NLTK、BERT等）进行文本预处理和关键信息提取。2. 对于图片格式的简历，利用图像处理和文字识别技术（如OCR）进行文字信息的提取和结构化数据转换。使用OCR工具（如Tesseract、Google Cloud Vision等）处理图片格式简历的文字识别。3. 根据具体需求，选择合适的简历处理方式：第一种解决方式（OCR处理）或第二种解决方式（直接文本处理）。难点：1. 准确识别简历中的关键信息：不同简历的结构和表述方式可能存在差异，使得关键信息的提取变得具有挑战性。2. 处理多样化的简历结构：简历的格式和内容可能有很多变化，需要设计灵活的算法以适应不同的简历结构。3. 提高文字识别准确性：针对图片格式简历，OCR技术可能会受到排版、字体、背景噪声等因素的影响，导致识别准确性下降。4. 文本处理中的乱序问题：对PDF和Docx文件直接提取文本时，可能会出现文本乱序的问题，这会影响后续关键信息的提取和处理。5. 整合不同来源的简历数据：由于简历可能来自于不同的来源和格式，需要设计合适的方法将各种格式的简历数据整合在一起。为了解决这些难点，可以尝试以下方法：1. 使用预训练的NLP模型（如BERT）以提高关键信息提取的准确性。2. 设计自适应的算法或规则，以处理不同结构和格式的简历。3. 对OCR技术进行优化和调整，以提高文字识别准确性。4. 在文本处理过程中，通过自然语言处理框架解决文本乱序问题，如利用关键词提取、实体识别等方法提取关键信息。

人岗匹配方面：

技术路线：数据预处理、特征工程、模型选择与训练、模型评估、模型优化以及结果解释与展示

难点：

1.在构建人岗匹配算法之前，需要对原始数据进行预处理，包括清洗异常或错误的数据，填充缺失值，处理极端值，将文本信息转换为数值表示……

2.人岗匹配中，设计合适的特征对于算法的性能至关重要。这可能包括提取教育背景、工作经验、技能等方面的特征，选择合适的权重就显得尤为重要了

3.选择合适的机器学习模型，如逻辑回归、支持向量机、决策树、随机森林、神经网络等。难点在于找到适用于人岗匹配问题的模型，以及调整模型参数以获得较好的性能。

四、成员分工

毛子昊：前端

杨力闻，林诰人：后端

牛锴鹏：算法