智能简历解析系统

概要设计说明书

2023年  7月

目录

[1 前言 3](#_Toc140398418)

[1.1 文档目的 3](#_Toc140398419)

[1.2 背景 3](#_Toc140398420)

[1.3 文档范围 3](#_Toc140398421)

[1.4 参考文档 3](#_Toc140398422)

[2 总体设计 3](#_Toc140398423)

[2.1 系统描述 3](#_Toc140398424)

[2.1.1 系统概述 3](#_Toc140398425)

[2.1.2 运行环境 4](#_Toc140398426)

[2.1.3硬件配置 4](#_Toc140398427)

[2.2 总体设计说明 5](#_Toc140398428)

[2.2.1 基本设计概述 5](#_Toc140398429)

[2.2.2 设计思想 5](#_Toc140398430)

[2.2.3 系统总体结构 5](#_Toc140398431)

[2.2.4 处理流程 6](#_Toc140398432)

[3 接口设计 7](#_Toc140398433)

[3.1 外部接口 7](#_Toc140398434)

[3.1.1 客户端接口 7](#_Toc140398435)

[3.1.2 PaddleNLP接口 8](#_Toc140398436)

[3.2 内部接口 8](#_Toc140398437)

[4 系统结构设计 8](#_Toc140398438)

[4.1 客户端模块 8](#_Toc140398439)

[4.1.1 简历信息录入 8](#_Toc140398440)

[4.1.2 简历信息统计与展示 9](#_Toc140398441)

[4.2 解析模块 12](#_Toc140398442)

[4.2.1 自动提取简历关键信息 12](#_Toc140398443)

[4.2.2 人才画像构建与岗位匹配 12](#_Toc140398444)

[5 运行设计 13](#_Toc140398445)

[5.1 运行模块的组合 13](#_Toc140398446)

[5.2 运行控制 14](#_Toc140398447)

[5.3 运行时间 14](#_Toc140398448)

# 1 前言

## 1.1 文档目的

文档编写目的是为是为了提供一个详细的指南，帮助开发人员和其他相关人员了解该系统的设计和功能。该文档旨在解释系统的需求、设计、实现和测试，并提供有关系统的所有方面的详细信息。

## 1.2 背景

该项目的背景是，随着互联网技术的快速发展，招聘市场变得越来越复杂和竞争激烈。人力资源部门需要处理成千上万份求职者的简历，以找到最合适的候选人。然而，这个过程非常耗时和繁琐，需要大量的人力和时间。

为了解决这个问题，智能简历分析系统应运而生。该系统基于人工智能和自然语言处理技术，可以自动解析、归类和筛选海量的简历，从中选出最合适的候选人，为企业提供高效、准确、快速的人才招聘服务。

该系统的目标是提高人力资源部门的工作效率和准确性，减少招聘成本和时间成本，同时也为求职者提供更好的招聘体验和更多的就业机会。

## 1.3 文档范围

产品范围：根据《智能简历解析系统需求分析规格书》，该文档阐述软件产品的功能

## 1.4 参考文档

《智能简历解析系统需求分析规格书》

# 2 总体设计

## 2.1 系统描述

### 2.1.1 系统概述

本系统采用三层架构设计，分为展示层、业务层和组件层。

1.展示层：负责用户交互，包括简历上传、结果展示等操作，使用pyqt实现。

2.业务层：处理主要业务逻辑，包括简历解析、人才画像构建等，使用PaddleNLP实现。

3.组件层：负责提供实现业务逻辑所需要的功能模块，包括简历解析模块，人才画像构建模块，统计分析模块，可视化模块等。

### 2.1.2 运行环境

#### 2.1.2.1操作系统：

Windows 7及以上版本（32位或64位）

Ubuntu 16.04及以上版本

macOS 10.12及以上版本

#### 2.1.2.2所需库：

Python：3.10.11

LAC==2.1.2

pypinyin==0.49.0

pymupdf==1.22.5

paddlepaddle==2.4.2

paddleocr==2.6.1.3

bs4==0.0.1

ordered\_set==4.1.0

pdf2image==1.16.3

pyqt6==6.4.2

qt-material==2.14

### 2.1.3硬件配置

#### 2.1.3.1最低硬件配置：

处理器: Intel Core 2 Duo E5200

内存: 4 GB RAM

显卡:不需要

#### 2.1.3.2推荐硬件配置：

处理器: Intel Core i5

内存: 8 GB RAM

显卡: GeForce GTX 560

## 2.2 总体设计说明

### 2.2.1 基本设计概述

从《需求规格说明书》出发，根据需求分析阶段确定的功能设计软件系统的整体结构、划分功能模块、确定每个模块的实现算法以及编写具体的代码，形成软件的具体设计方案。

### 2.2.2 设计思想

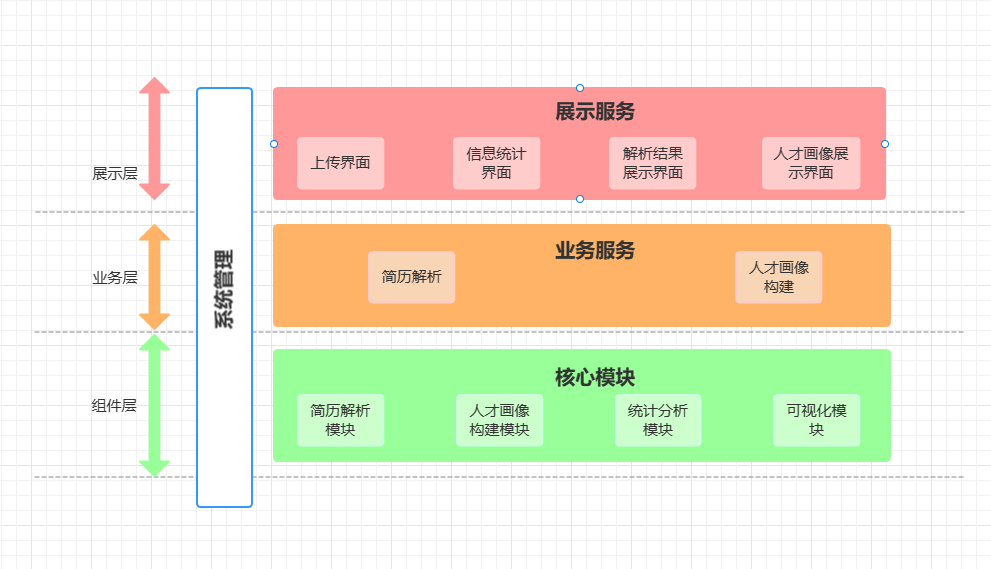
本概要设计主要由结构设计、接口设计、全局数据结构设计及其他方面设计组成。需遵循软抽象、模块化、信息隐蔽及模块独立性等特征。模块独立性从两个方面度量：

(1)内聚性：偶然内聚、逻辑内聚、时间内聚、过程内聚、通信内聚、顺序内聚、功能内聚。

(2)耦合性：内容耦合、公共耦合、外部耦合、控制耦合、标记耦合、数据耦合、非直接耦合

### 2.2.3 系统总体结构

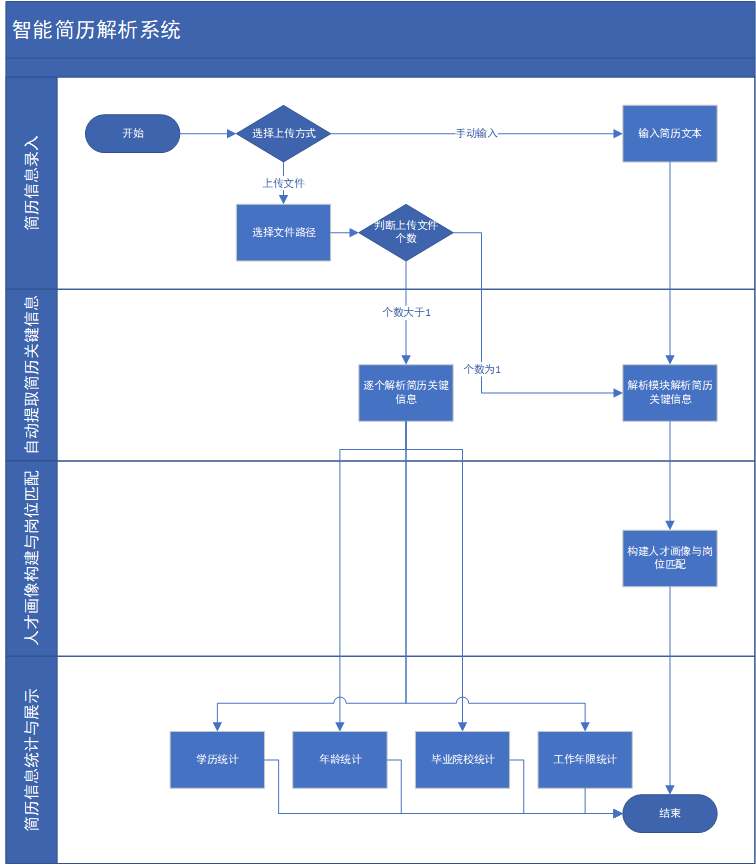
系统总体架构图（图2-1）



系统总体结构图（图2-1）

### 2.2.4 处理流程

系统流程处理图（图2-2）



系统流程处理图（图2-2）

# 3 接口设计

## 3.1 外部接口

### 3.1.1 客户端接口

客户端使用PyQt框架进行开发，其接口如下：

1.上传简历文件：提供一个上传简历文件的按钮，用户可以点击按钮选择要上传的简历文件。

2.展示简历信息：在界面上展示简历信息，包括基本信息、教育背景、工作经历、技能特长等方面的内容。

3.构建人才画像：提供一个按钮，用户可以点击按钮构建人才画像，并在界面上展示人才画像信息。

### 3.1.2 PaddleNLP接口

本系统使用PaddleNLP进行简历解析和人才画像的构建，其接口如下：

1.简历解析接口：调用PaddleNLP提供的简历解析模型进行简历解析。

2.人才画像构建接口：调用PaddleNLP提供的机器学习和数据挖掘技术，构建求职者的人才画像。

## 3.2 内部接口

在MVC架构中，各层之间的交互主要通过接口完成，以下是主要接口的设计：

1.简历上传接口：用户通过此接口上传简历文件，接口接收文件路径参数，返回上传结果。

2.简历解析接口：接收简历文件路径参数，返回解析结果。

3.人才画像构建接口：接收解析结果参数，返回人才画像。

4.结果展示接口：接收解析和构建的结果参数，返回展示结果。

# 4 系统结构设计

## 4.1 客户端模块

### 4.1.1 简历信息录入

1. 界面原型

文件上传简历原型界面（图4-1-1）



文件上传简历原型界面（图4-1-1）

手动输入简历原型界面（图4-1-2）



手动输入简历原型界面（图4-1-2）

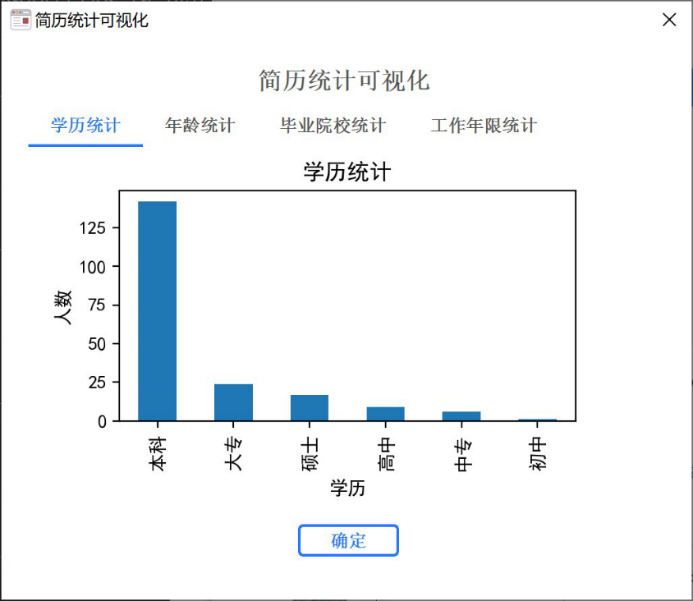
2. 实现功能

用户通过此模块上传简历文件，支持多种文件格式，如.docx、pdf、png等，同时支持手动输入简历信息。

### 4.1.2 简历信息统计与展示

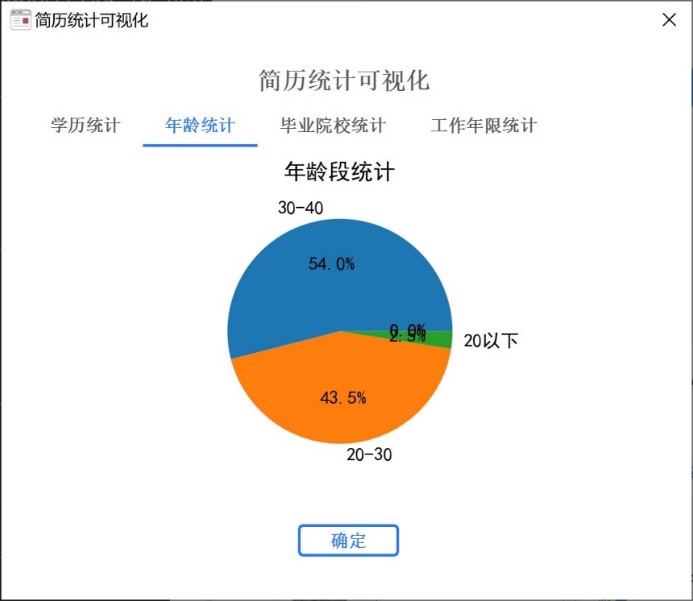
1. 界面原型

学历统计原型界面（图4-2-1）



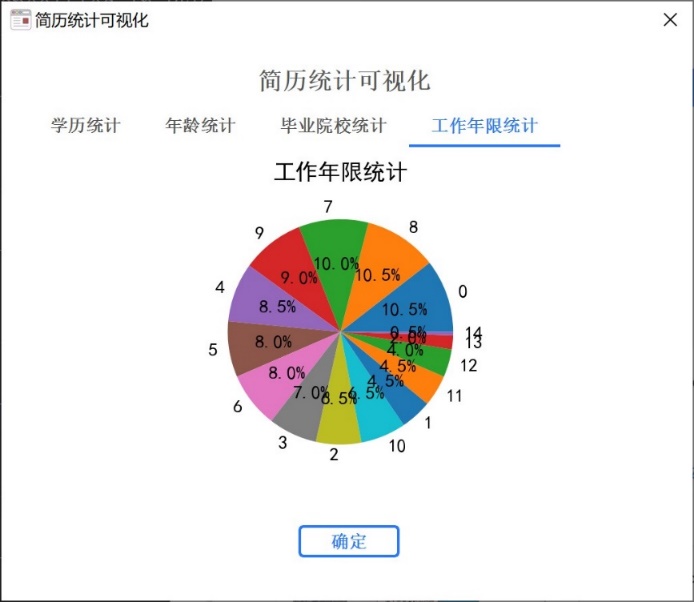
学历统计原型界面（图4-2-1）

学历统计原型界面（图4-2-1）



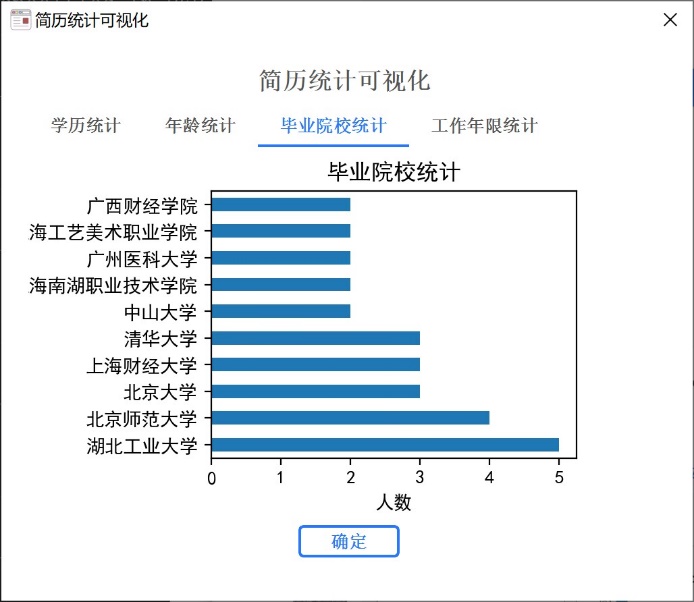
年龄统计原型界面（图4-2-2）

工作年限统计原型界面（图4-2-3）



工作年限统计统计原型界面（图4-2-3）

毕业院校统计原型界面（图4-2-4）



毕业院校统计原型界面（图4-2-4）

2. 实现功能

本模块将多份简历的统计数据可视化，以图表的形式呈现给用户。

支持以柱状图的形式展示学历统计数据和毕业院校统计数据，以饼状图的形式展示年龄统计数据与工作年限统计数据。

## 4.2 解析模块

### 4.2.1 自动提取简历关键信息

1. 界面原型

简历解析结果原型界面（图4-2-5）



简历解析结果原型界面（图4-2-5）

2. 功能说明

本模块通过百度飞浆平台的ocr模型，提取上传文件的文本信息，并通过nlp模型自动识别提取简历文本当中的姓名、出生年月、年龄、政治面貌、学历、电话、毕业院校、工作年限等信息。

### 4.2.2 人才画像构建与岗位匹配

1. 界面原型

人才画像构建与岗位匹配原型界面（图4-2-6）



人才画像构建与岗位匹配原型界面（图4-2-6）

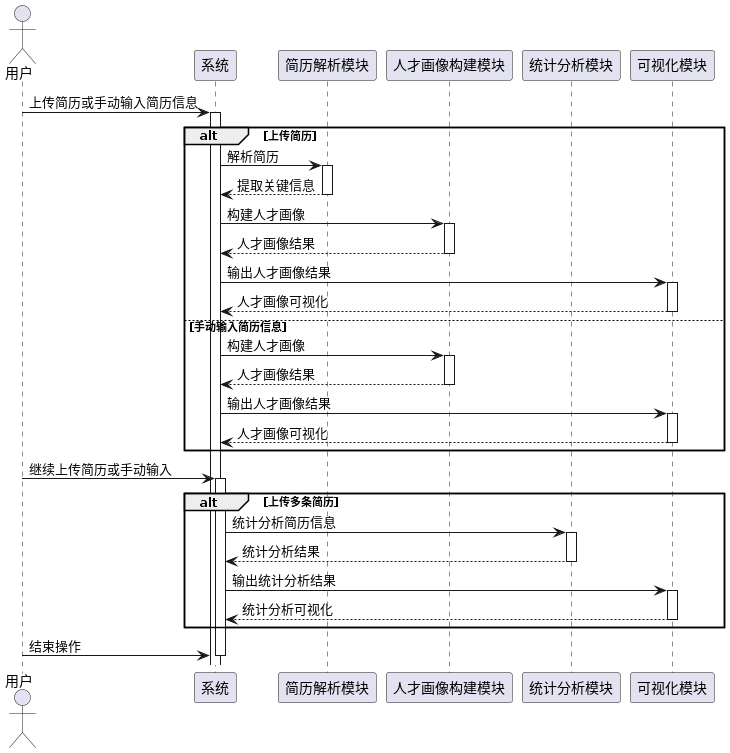
2. 功能说明

本模块根据自动提取到的简历的关键信息，为应聘者标记不同的选手标签，与预期薪资，再根据标签匹配合适的岗位。

# 5 运行设计

## 5.1 运行模块的组合

运行模块时，系统时序图（图5--1）



系统时序图（图5-1）

## 5.2 运行控制

运行控制将严格按照各模块间函数调用关系来实现。在各事务中心模块中，需对运行控制进行正确的判断，选择正确的运行控制路径。

## 5.3 运行时间

本系统的简历解析过程对性能要求较高，因此在进行多个简历解析任务时需要使用移动云云上AI的算力，以缩短程序的运行实践，提高程序的运行效率。