基于互联网大数据的计算机相关专业招聘分析与应用（系统说明文档）

团队成员：简雁，武盛弘，郗华鹏，李静，徐静如

（所在单位：曲阜师范大学软件学院，所在地区：山东省济宁市曲阜市，邮政编码：273100）

1. 系统的运行环境

该系统为B/S三层结构，它的运行环境分网页端、应用服务两部分。

以下是系统的软件环境。

（1）网页端

谷歌浏览器

版本：6.6及以上版本

（2）应用服务器端

本地运行

1. 使用方法

将压缩包解压后导入pycharm，安装程序所需python第三方库后，在pycharm里运行func函数，之后打开谷歌浏览器输入localhost:5000进行访问

1. 测试方法

在网站上上传图片，得出的人数与图片上的人数进行对比

* 1. 总体结构处理流图

****

s主要使用技术

jieba分词

输入文本，jieba分词可以产生对应文本的分词。jieba分词自带的posseg.cut可产生(词，词频)的元组，同时jieba分词的jieba.cut可以根据指定词性进行分词筛选，两部分结合使用Counter.most\_commmon(x)可产生指定词性的出现次数前x的分词结果的列表

gensim库

通过gensim库自带的Word2vec可以通过读取文件中的内容使用神经网络训练获取与给定内容最为相近的词。通过输出匹配度数值大的结果可按顺序显示匹配内容。

词云

通过创建wordcloud对象，传入jieba分词分好的词库可产生给定图片为背景的词云

sql

通过sql可查询给定要求的数据库的内容

正则表达式

通过正则表达式可以筛选替换文本中的内容