



闽江学院

本科毕业论文（设计）

题 目	基于大数据技术的招聘网站信息数据之分析与应用
学 生 姓 名	陈晞玟
学 号	3187103111
学 院	数学与数据科学学院（软件学院）
年 级	2018 级软件工程（闽台合作）一班
专 业	软件工程
指 导 教 师	曹永忠
职 称	助理教授
完 成 日 期	2022 年 4 月 29 日

闽江学院毕业论文（设计）诚信声明书

本人郑重声明：

兹提交的毕业论文（设计）《基于大数据技术的招聘网站信息数据之分析与应用》，是本人在指导老师曹永忠的指导下独立研究、撰写的成果；论文（设计）未剽窃、抄袭他人的学术观点、思想和成果，未篡改研究数据，论文（设计）中所引用的文字、研究成果均已在论文（设计）中以明确的方式标明；在毕业论文（设计）工作过程中，本人恪守学术规范，遵守学校有关规定，依法享有和承担由此论文（设计）产生的权利和责任。

声明人（签名）：陈晔致

2022年4月29日

摘 要

随着信息时代的到来，人们对工作的需求越来越大。盲目参加招聘会需要时间和精力。招聘信息数据分析系统，以多个热门城市的相关岗位信息为数据源，分析我国多个岗位的地域、学历、薪资需求，呈现调研结果的可视化关系，分析岗位的招聘需求现状。方便企业和求职者节省人力和财力。

对于求职者来说，真实有效的数据是必要的。但通过传统的方式，是非常昂贵且消耗人力的。所以，我们运用大数据技术获取和分析招聘网站数据，根据用户所选择的个性和倾向，进行工作需求分析。

本论文主要阐述了系统的研发流程:资料来源、需求分析、数据库设计、系统实现。对整个系统结构、功能设计的基本过程和各个功能模块作了全面的介绍和比较具体的剖析和研究。

关键词：大数据；招聘网站信息；数据库

Abstract

With the advent of the information age, there is an increasing demand for work. It takes time and energy to attend job fairs blindly. The recruitment information data analysis system takes the relevant position information of many popular cities as the data source, analyzes the region, education and salary demand of many positions in China, presents the visual relationship of the research results, and analyzes the current situation of the recruitment demand of positions. Convenient for enterprises and job seekers to save manpower and financial resources.

For job seekers, real and effective data are necessary. But in the traditional way, it is very expensive and costly. Therefore, we use big data technology to obtain and analyze the data of recruitment websites, and analyze the job needs according to the chosen personality and tendency of users.

This paper mainly expounds the development process of the system: data source, demand analysis, database design and system implementation. The whole system structure, the basic process of functional design and the basic process of each functional module has made a comprehensive introduction and more specific analysis and research.

Key words: Big Data; Recruitment Website Information; Database

目 录

1 序论	4
1.1 研究背景	4
1.2 研究目的	4
1.3 论文架构图	6
2 应用原理和理论依据	7
2.1 大数据的运用	7
2.2 资料来源	8
2.3 资料列举	10
2.4 技术方案及措施	11
2.5 开发工具与运行环境	13
3 系统分析与设计	15
3.1 需求分析	15
3.2 系统结构设计	15
3.3 系统架构设计	16
4 数据库设计	17
4.1 逻辑结构设计	17
4.2 资料表栏位规格书	17
5 系统实现	21
5.1 系统模块介绍	21
5.2 系统页面展示	21
6 总结及未来展望	31
6.1 总结	31
6.2 未来展望	32
参考文献	33
致谢	34

附录	35
附录 1	35
附录 2	- 42 -
附录 3	- 43 -

基于大数据技术的招聘网站信息数据之分析与应用

陈晞玟

(闽江学院 数学与数据科学学院 (软件学院), 福建 福州 350108)

1 序论

1.1 研究背景

在手机网络技术迅猛发展的今天, 公司与公司的竞争已由资本的争夺转变为对人力资源的争夺。在知识经济的今天, 人才已经成为了是公司发展的重要因素, 提高员工的素质和智能感知能力, 正确且精确地使用招聘数据信息分析技术是公司的人力资源部和管理者应该具备的基本技能; 这是政策制定者需要考虑的问题。要想在人才资源上占有绝对的上风, 维持公司的长期发展势头和竞争能力, 就必须充分认识到, 拥有优秀的员工是公司发展的关键。在目前的求职形势下, 由于应聘工作的困难和业内的激烈, 公司期望能够在招聘服务上有所改进, 以提升公司的招募效率, 提升顾客的满意程度, 从而提升公司的工作效率, 从而在整个国家的人才市场中占据更大的份额^[19]。文章围绕针对前景无忧的求职招聘业务, 从招聘的内容和招聘过程中发现的问题进行了剖析, 并就如何改进工作进行了探讨。对未来的发展进行了深入的研究; 针对当前大学生、社工们所关心的就业问题, 基于前程无忧多年的工作实践与调查, 针对当前问题, 建议改进措施, 改变宣传流程策略, 提供个性化服务模式, 强化总结反馈。本文主要是针对前程无忧招聘网站服务提出改善, 目的是通过对现有招聘服务内容上进行科学、系统地优化后, 能够为企业招聘提供精确且有效的服务, 展现出招聘的规范性、专业性^[3]。

大数据时代的到来给中国经济社会的发展带来了巨大的变化。问题解决的最关键就在于, 求职招聘活动组织者和企业招聘部门负责人应该怎样去寻找一个合适于自己发展的招聘岗位, 怎样去录用一个合格优秀的招聘人员来去补充招聘企业人员的岗位缺口^[20]。

1.2 研究目的

设计这本招聘系统也就是为了能够顺应了这个大趋势, 借助这种信息平台,

可以做到不用每天出门就已经能够随时掌握各个招聘信息网站上的最新招聘职位信息数据的进行详细的分析，给社会大众求职带来方便。

本文是对互联网上地球数据处理和处理的研究。当今的企业在不断发展、变革和创新，企业对人才的标准要求很高，人才的素质在不断提高，每个人的工作计划都是不一样的。因此，我们必须提高招聘和求职的效率，让企业和人才更好地为社会服务。通过自主研发的招聘信息数据分析处理系统，可以体验到优质的信息、科学的统计和快速查找适宜的工作，提高工作效率，提升服务质量。

该系统为用户提供从招聘需求分析到招聘查找再到应聘报到全方位的一站式解决方案，提供各项相应的服务，并按照用户的选择的个人条件和倾向选择查找适合的岗位信息，为个人求职，个人发展提供可持续发展动力的全方位服务，利用网络的便捷为各个公司寻找符合其需求的各种专业人才同时也为各个行业、各个地区的人才寻找到一份能够充分体现自身价值的工作。此系统具有低成本、大容量、速度快和强调个性化服务的优势，允许更加灵活的方式，提供更丰富的信息资源。

1.3 论文架构图

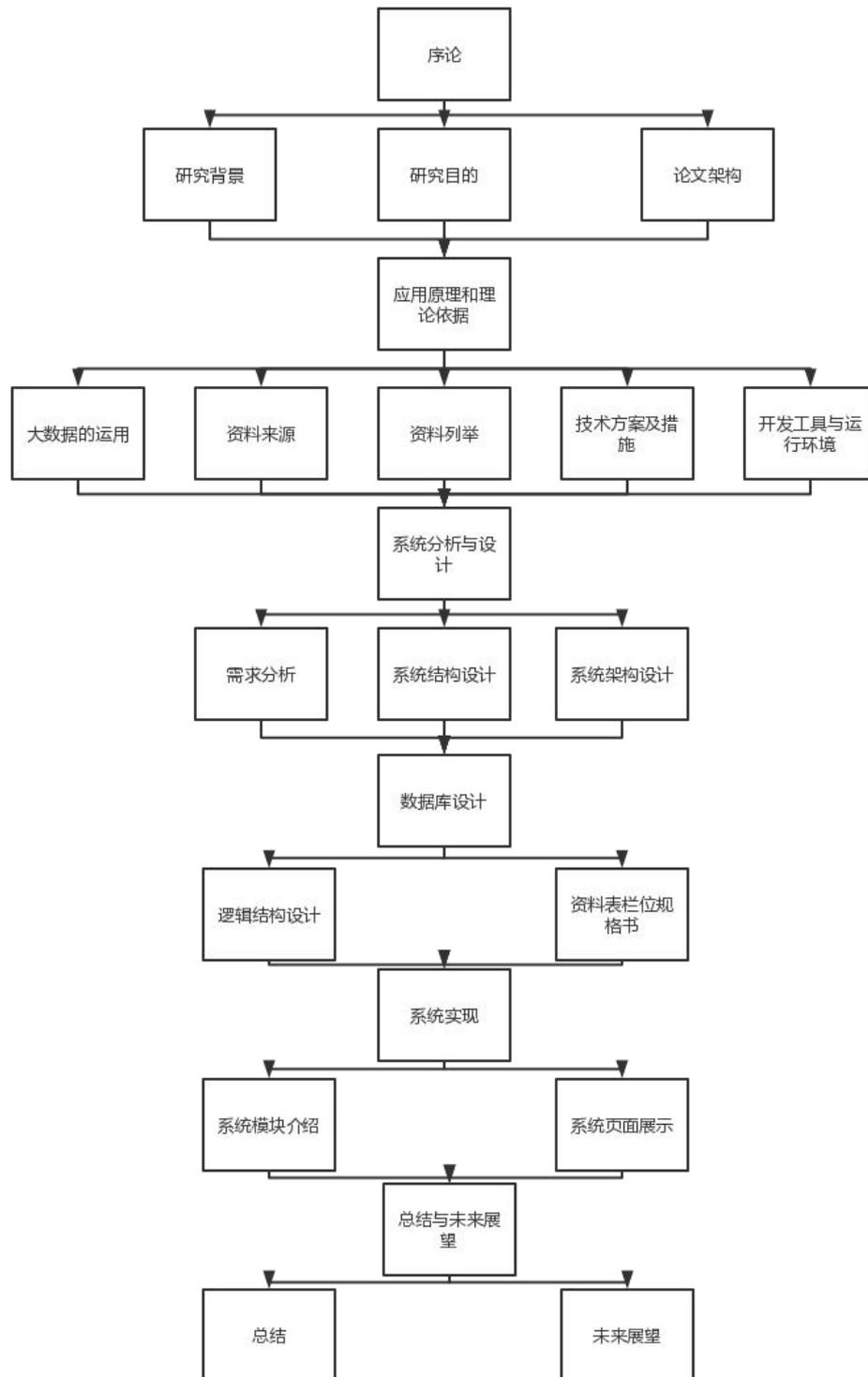


图 1-1 论文架构图

2 应用原理和理论依据

本章节主要介绍了大数据的特点、信息处理技术的关键价值，此系统的资料来源和此次运用到的技术方案及措施。

2.1 大数据的运用

大数据是指超越传统信息数据库系统信息处理技术能力的信息数据。它的统计规模和传输速度都很高,或者说其架构不符合原有的数据库系统。

大数据的特点:

1. 价值化 **Value**: 大量数据不相关的资讯统计,对未来变化趋势和发展模型的可预见解析,以及深入而复杂性的数据挖掘(机器学习、人工智能与传统商业智慧)。
2. 多样化 **Variety**: 信息的异质性和多样性,多种不同表现形式(文字、图片、视讯、机械资料数据),无模型或模型不明确,语法或句子含义不连续。
3. 敏捷化 **Velocity**: 即时解析而非批量解析,数据信息进入、处理过程和丢失,很快就会出现效果。
4. 真实性 **Veracity**: 信息的信息内容与现实世界的发生密切相关。信息数据的研究是从海量网络信息数据中获取到能够解释和预测真实发生事件的过程^[1]。
5. 大量化 **Volume**: 非结构化数据的规模和增长,占总已知数据的 80%~90%,比结构化数据快 10~50 倍,比传统数据仓库快 10~50 倍^[2]。

数据分析的关键价值之一是通过现有数据分析预测未来数据的发展趋势,为行业发展提供更好的预测数据。预测分析的主要目的是确定数据属性,创建科学的数据模型,引入新数据,获得新的预测结果,以供开发过程中参考。

信息技术也是大数据处理信息技术的重要路径。

1. 云计算的普及。大数据对计算机计算能力和存储技术提出了新的更高的要求,而云计算的诞生解决了这一重要的技术瓶颈。云计算将成为一种存储工具和应用载体,提供超越极限的强大计算能力和存储空间,满足大数据的技术需求,带动大数据的兴起和繁荣。云计算又可以分为“存储云”和“计算云”。总的来说,云存储是一种以数据存储和管理为核心的云计算系统^[4-5]。

2. 信息处理技术创新。为了增加信息和数据的价值，需要提高计算效率和方法，已经开发了许多方法来提高大数据的处理效率，使用 VC 的 OO4O 中包含的 Oracle C++Class Library。在 ASP.Net 中；SOAP 协议改进大量数据的传输性能。
3. 信息分析方法的创新。人们在应用和分析大数据的过程中，针对新情况、新问题开发新技术，不断加快数据分析处理技术的发展和进步。目前，大数据分析的主要方法是全数据分析和相关性分析。

这两种方法超越了传统取样分析方法的要求和对样品准确性的依赖，分析结果更加全面准确，使复杂数据更加灵活高效^[12]。

因此，当前在大数据的时代背景下，应该要不停的完善优化和提高数据信息分析处理能力，而从业者必须抓住这个机会。积极加强各种数据信息分析处理技术的优化和创新，这样才能以更好的姿态应对大数据时代的新挑战。

2.2 资料来源

本次设计主要是以前程无忧网站的多个热门城市的相关招聘信息为数据来源，基于 python 语言实现爬虫框架获取数据，模拟浏览器对网站发起 HTTP 请求，得到各个地区招聘的基本数据，岗位需求的技能、岗位薪资、岗位位置分布城市等信息的日志文件，以 json 对象存储^[16-17]。用大数据分析框架 spark 或 pandas¹对数据进行空值的填充和敏感信息的脱敏加密，将半结构化数据处理为结构化数据，储存进入到数据库建表，此时数据杂乱，所以需要数据清洗，对数据进行多维度分析，包括岗位数量和薪资水平、工作经验和学历要求、岗位关键字和技能要求、经验与薪资的关系等维度，提高明细表的易用性，将分析的结果建立 ads 层²，存放在数据库中^[15]。页面采用 Cascading Style Sheets³+HyperText Markup Language⁴+JavaScript⁵编写，采用前端的 vue⁶框架进行最后的可视化展示。

以学历为例，学历在一定程度上反映了高校学生的专业知识水平与工作基本

¹ pandas 是基于 NumPy 的一种工具，该工具是为解决数据分析任务而创建的。Pandas 纳入了大量库和一些标准的数据模型，提供了高效地操作大型数据集所需的工具。

² ads 层主要就是最后需要分析的数据,可以从各种纬度,这就需要根据业务来判断最后选择哪一些纬度来进行分析,

³ 层叠样式表(英文全称: Cascading Style Sheets)是一种用来表现 HTML (标准通用标记语言的一个应用)或 XML (标准通用标记语言的一个子集)等文件样式的计算机语言。

⁴ HTML 的全称为超文本标记语言,是一种标记语言。它包括一系列标签。通过这些标签可以将网络上的文档格式统一,使分散的 Internet 资源连接为一个逻辑整体。HTML 文本是由 HTML 命令组成的描述性文本,HTML 命令可以说明文字,图形、动画、声音、表格、链接等。

⁵ JavaScript (简称“JS”) 是一种具有函数优先的轻量级,解释型或即时编译型的编程语言。

⁶ VUE 是 iOS 和 Android 平台上的一款 Vlog 社区与编辑工具。

能力，是企业招聘中的重要参考指标。主要是将学历划分为：中技以下、中技（中专）、高中、大专、本科、硕士、博士和不限，对其分组，以招聘公司数（同一公司数据岗位对学历的要求是一致的）进行聚合，求得各个学历阶段的公司数量，如表 2-2 所示，公司的数量就是该学历下，求职者学历的要求占比。分析数据如表 2-1 所示：

表 2-1 抓取文献资料程序简码

```
def diploma(self):
    job_diploma = self.get_source() \
        .groupby("diploma") \
        .agg(F.countDistinct(F.col("company")).alias("user")) \
        .orderBy(F.col("user").desc())
    MySQLUtils().sink_to_mysql(job_diploma,table='job_diploma')

    return job_diploma
```

学历	公司数量
本科	2847
大专	1248
硕士	303
博士	10
不限	201
中技	44
高中	28

表 2-2 学历岗位数量

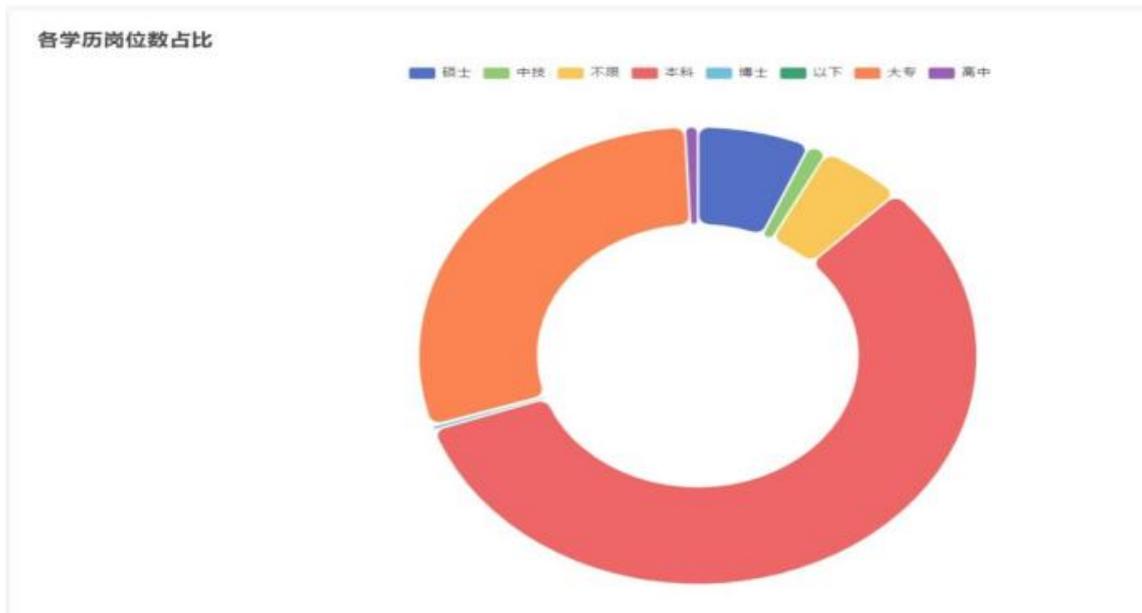


图 2-3 学历岗位数量占比

据统计，数据类岗位中超过 90%的招聘信息对学历提出明确的要求，如图 2-3 所示，其中，数据分析师岗位对教育背景要求更为明确，主要以本科学历为主，其次是大专学历，而对硕士，对学历要求没有限制的岗位数据只占到整体的约为 5%，几乎没有公司招聘博士学历。数据分析师对于业务的理解和数据敏感，在学历上可以根据数据分析的经验和技能来提升自己。所以大专学历及以下仍能够找聚到近 30%的比例。

数据分析专家对学历的相对要求会高很多，求以硕士和本科学历为主，博士学历为辅，且该岗位对硕士学历的需求量远远超过了本科生。相比之下，数据分析师对学历的要求明确且工作的专业性比较强，对于计算机相关专业的高学历相关专业人才的需求量更大。而对于数据类岗位招聘来说，学历要求越高，能达到就业的公司岗位专业要求门槛范围越窄，上岗后工作内容专业性也比较强。

我们通过对数据进行清洗和数据可视化等方法对所获取到的数据进行分析与应用。根据大数据理论，以多个热门城市的相关招聘信息为数据源，分析我国多个地区岗位的地域分布、热点分布、薪资水平现状，提出研究成果，并展现结果的可视化关系。在我国发现了当前各企业的招聘需求^[6]。

2.3 资料列举

本系统资料来源于前程无忧网站的多个热门城市的相关招聘信息，下面列举部分数据,如表 2-4 所示。

表 2-4 部分数据列表

	地区数（占比）	公司数（占比）	数据数（占比）
上海	1（4.3%）	1713（12.45%）	331（7.93%）
北京	1（4.3%）	1433（10.42%）	307（7.36%）
青岛	1（4.3%）	839（6.1%）	398（7.14%）
深圳	1（4.3%）	1674（12.74%）	376（9.01%）
总数	23（100%）	11363（100%）	41126（100%）

2.4 技术方案及措施

此系统主要用到数据采集、数据预处理、数据清洗、数据可视化对获取到的数据进行分析。

1. 数据采集：运用 Python 语言编写代码，对对应模块发起 request 请求，爬取数据，将爬取的元数据数据，存储在磁盘上，获得多个分类不同的 log 文件。如图 2-5 所示：



图 2-5 数据采集流程

2. 数据预处理：本次设计主要是以前程无忧网站为数据来源，大数据框架 spark 对日志的数据进行预处理，脱敏加密不合法数据，对当中的缺失值或空值进行填充或者过滤。将半结构化数据处理为结构化数据，写入 mysql 中，建立 ods

- 层数据⁷，建立宽表。部分市场报告给出的 csv 数据⁸，利用 python 库 pandas，对数据进行预处理，写入 mysql 与宽表⁹关联使用。
3. 数据清洗：采用三范式建模思想¹⁰，Spark 抽取 ods 层数据，创建 dwd 层¹¹数据，运用维度降级方法将维度表收缩为事实表，易于使用。采用维度分析，建立维度表，编写 python 程序，运行 spark 进行数据分析，将分析的结果建立 ads 层，存放在数据库中。
 4. 数据可视化：如图 2-6 所示，将存在于数据库中 ads 层的个性化的统计指标数据进行前端展示。主要使用 spring 框架实现后端代码，在 entity¹²建立与 ads 层相对应的实体类，在 daoimpl 层实现 dao 接口¹³，将数据库中的数据给前端可视化，采用前端的 vue 框架进行最后的可视化展示，以助于更直观地了解大量的复杂数据。

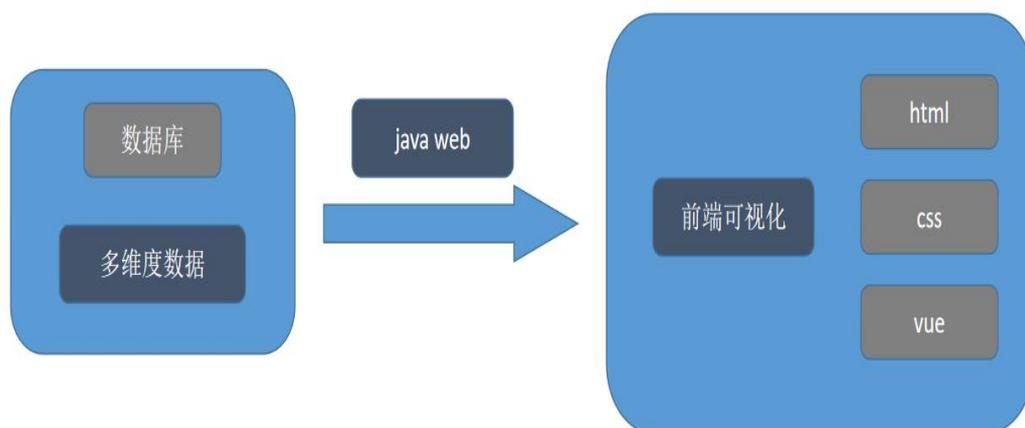


图 2-6 数据可视化流程

通过对数据进行清洗和数据可视化等方法对所获取到的数据进行分析：如图 2-7 所示，以多个热门城市的相关招聘信息为数据源，分析我国多个地区岗位的地

⁷ ODS 全称是 Operational Data Store,操作数据存储.“面向主题的”,数据运营层,也叫 ODS 层,是最接近数据源中数据的一层,数据源中的数据,经过抽取、洗净、传输,也就说传说中的 ETL 之后,装入本层。

⁸ csv 是【comma-separated values】的缩写,代表一种包含数据列表的纯文本文件格式,这些 csv 文件通常用于在不同应用程序之间交换数据。

⁹ 宽表从字面意义上讲就是字段比较多的数据库表。通常是指业务主题相关的指标、维度、属性关联在一起的一张数据库表。

¹⁰ 所有表中的数据都为原子数据,不可再分、所有表中的所有字段都必须依赖主关键字、所有表中的非主关键词之间不能函数依赖关系。

¹¹ DWD(Data WareHouse Detail)数据明细层,主要是将从业务数据库中同步过来的 ODS 层数据进行清洗和整合成相应的事实表。

¹² entity 是实体类的意思,主要功能是存储数据,一般的结构与数据库表结构相同。

¹³ DAO(数据访问对象)是一种应用程序编程接口(API),存在于微软的 Visual Basic 中,它允许程序员请求对微软的 Access 数据库的访问.DAO 是微软的第一个面向对象的数据库接口。

域分布、热点分布、薪资水平现状，展现结果的可视化关系^[7]，在我国发现了当前各企业的招聘需求。

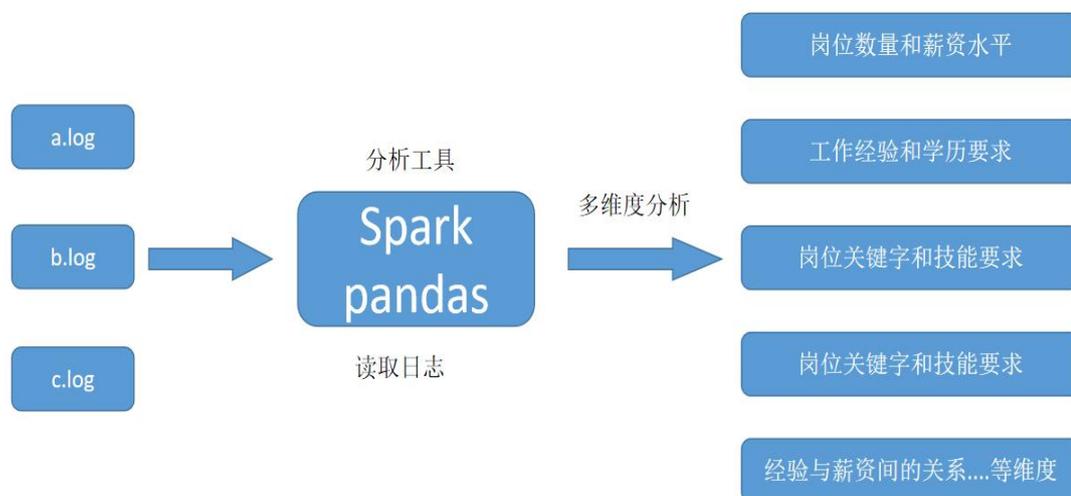


图 2-7 数据分析流程

2.5 开发工具与运行环境

这个系统是以 `pycharm 2020.3`¹⁴、`idea 2020.3`¹⁵为开发工具。经过多年的发展，这两种技术的广泛应用和稳定的性能为表示层技术奠定了坚实的基础。

数据采集使用爬虫技术。Python 利用从网站上爬取的大量信息进行数据分析和研究。通过构建对收集到的数据进行可视化分析的框架，求职者在浏览信息时可以更深入地了解岗位信息，有效判断岗位信息是否合适，进而提高招聘效率^{[13][14]}。

使用 `pyspark 2.4.7`¹⁶和 `Pandas`¹⁷进行数据分析。`pyspark` 是 Spark 对 Python API 接口。您可以在 Python 环境中调用 `pyspark` 模块运行 Spark，完成大数据框架下的数据分析和挖掘^[11]。`pandas` 是一个数据处理工具，集成了 `numpy` 和 `matplotlib`，具有便捷的数据处理和文件读取能力。

数据库采用 `mysql 5.7`¹⁸。SQL 是专门为数据库设计的操作命令集。它是一种功能十分齐全的数据库语言，可以轻松执行数据控制、数据操作和数据定义功能。

后端使用 `Javaweb`¹⁹：`Spring Boot 2.5.8`²⁰、`jdk 1.8`、`mybatis-plus 3.5.9`²¹。使用

¹⁴ 下载参考网址：<https://www.jetbrains.com/pycharm/>

¹⁵ 下载参考网址：<https://www.mpyit.com/intellijidea2020.html>

¹⁶ 下载参考网址：<https://spark.apache.org/docs/2.4.7/api/python/index.html>

¹⁷ 下载参考网址：<https://pandas.pydata.org/>

¹⁸ 下载参考网址：<https://dev.mysql.com/doc/relnotes/mysql/5.7/en/>

¹⁹ 下载参考网址：https://www.java.com/zh_CN/

javaweb 简单、直接且易于使用，编码、输入和输出都在同一个页面上，轻松查看所需内容。代码使用 Spring Boot²²架构和 B/S 结构²³，通过 Web 界面提交代码和输入，运行发送请求到服务器，服务器通过中间件编译并执行代码并将结果返回给浏览器，浏览器处理并显示它^[18]。

前端使用 vue 3。vue 3 不仅全面支持 TypeScript²⁴语法，还对生命周期进行了改善和剔除，代码简练有条理，在编码的过程中帮助我们检查类型化等问题。

图表用到 echarts²⁵。echarts 是一个提供了一些直观，易用的交互方式以便于对展示数据进行挖掘、提取、修正或整合，拥有互动图形用户界面的深度数据可视化工具。

²⁰ 下载参考网址：<https://spring.io/blog/2021/12/21/spring-boot-2-5-8-available-now>

²¹ 下载参考网址：<https://mybatis.org/mybatis-3/>

²² Spring Boot 是由 Pivotal 团队提供的全新框架，其设计目的是用来简化新 Spring 应用的初始搭建以及开发过程。

²³ B/S 结构是 WEB 兴起后的一种网络结构模式，WEB 浏览器是客户端最主要的应用软件。

²⁴ TypeScript 是微软开发的一个开源的编程语言，通过在 JavaScript 的基础上添加静态类型定义构建而成。TypeScript 通过 TypeScript 编译器或 Babel 转译为 JavaScript 代码，可运行在任何浏览器，任何操作系统。

²⁵ 下载参考网址：<https://echarts.apache.org/zh/index.html>

3 系统分析与设计

本章节主要介绍了系统分析与设计，其中包括需求分析，系统结构设计与系统架构设计，可以全面了解本系统的工作情况、系统处理过程和系统架构。

3.1 需求分析

需求分析任务通过探索现实世界中正在处理的对象，并阐明不同之处，从而全面了解原始系统的工作概况。了解用户需求并据此确定新功能。新系统需要仔细考虑未来的扩展和变化，而数据库不能仅仅根据当前应用程序的需求来设计。

与此同时，在线求职也变得越来越复杂和多样化。随着计算机的普及，网上收集的人力资源信息越来越多，招聘形式也越来越多样化。它分析和处理不同的岗位、地区、薪资和标签，使处理求职者信息变得更加方便，增强可靠性、安全性、完整性和易用性。

求职者和职位列表数据分析旨在帮助求职者和公司分析招聘信息，预测未来发展趋势，为行业发展提供更好的预测数据。此系统不仅能选择地区、工作经验和学历等动态数据生成分析柱状图，帮助求职者准确快速了解到现有的岗位信息现状，还能够提供职位列表，不仅能分析各个岗位的招聘数据，而且还能根据自身的学历、工作经验以及预期薪资搜寻适合自己的岗位，可以跳转岗位链接迅速了解详情。具有足够量的信息和快速查询选项的企业。随着科学技术的不断进步，计算机科学技术越来越成熟，可以显着提高招聘效率，减少求职者和招聘人员的工作量，为双方节省时间。

3.2 系统结构设计

本系统是求职招聘信息数据分析与运用系统。根据需求分析报告和可行性研究结果，全面了解系统处理的过程，可利用此系统流程来分析此系统的合理性。从总体上看，系统流程如下图 3-2 所示：

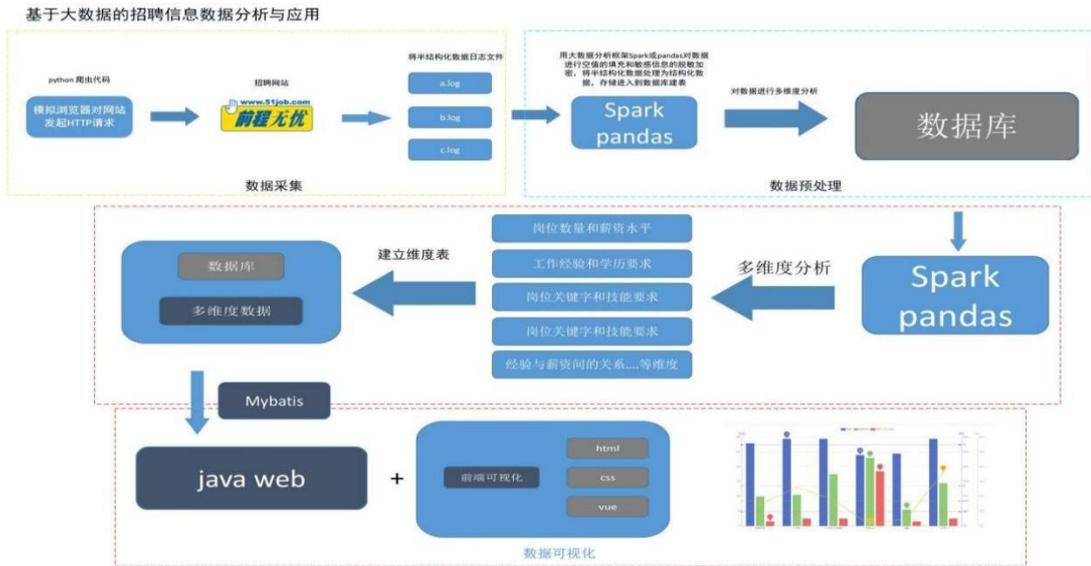


图 3-2 系统流程图

3.3 系统架构设计

本系统的架构设计主要分成了以数据报表、职位分析、地区分析、薪资分析、标签分析和用户管理为主的六大模块进行开发设计，而在此模块的基础上，又细分成了地区综合分析、城市薪资箱位图、平均薪资分析、岗位数量对比、薪资与岗位关系、薪资学历岗位关系、薪资经验岗位关系、行业分析、学历占比、工作经验和基本设置等子功能模块。系统的开发主要是为了让招聘信息处理工作更加规范化、系统化和细致化，提高处理信息的准确性和效率，提高做求职者和企业招聘的用户体验。其系统架构设计如图 3-3 所示：

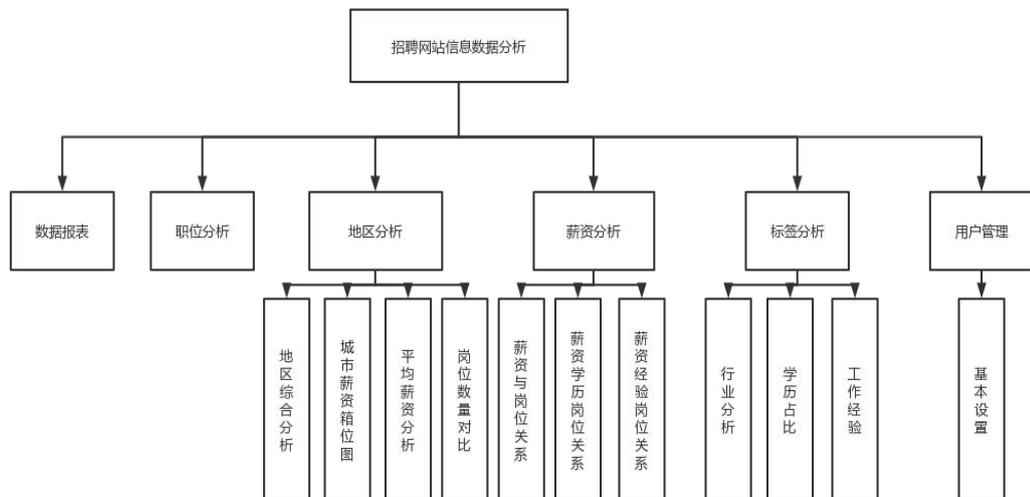


图 3-3 系统架构图

4 数据库设计

本章节主要介绍逻辑结构设计和资料表,可以通过图 5-1 表关系图了解到本系统的逻辑结构,根据系统需求分析和功能模块设计,绘制资料表栏位规格书,也可以参考附录 3,使用 SQL 叙述建立的 wind 资料表。

4.1 逻辑结构设计

逻辑结构的设计独立于任何数据模型。事实上,经常使用的数据库环境(SQL Server、Oracle、MySQL 等)已经证明了它的价值^[8]。由于目前使用的数据库基本是关系数据库,所以首先将其转换为关系模型,然后再转换为特定 DBMS 支持的数据模型,具体取决于该特定 DBMS 的特点和局限性,最后需要进行优化^{[9][10]}。表关系图如图 4-1 所示:

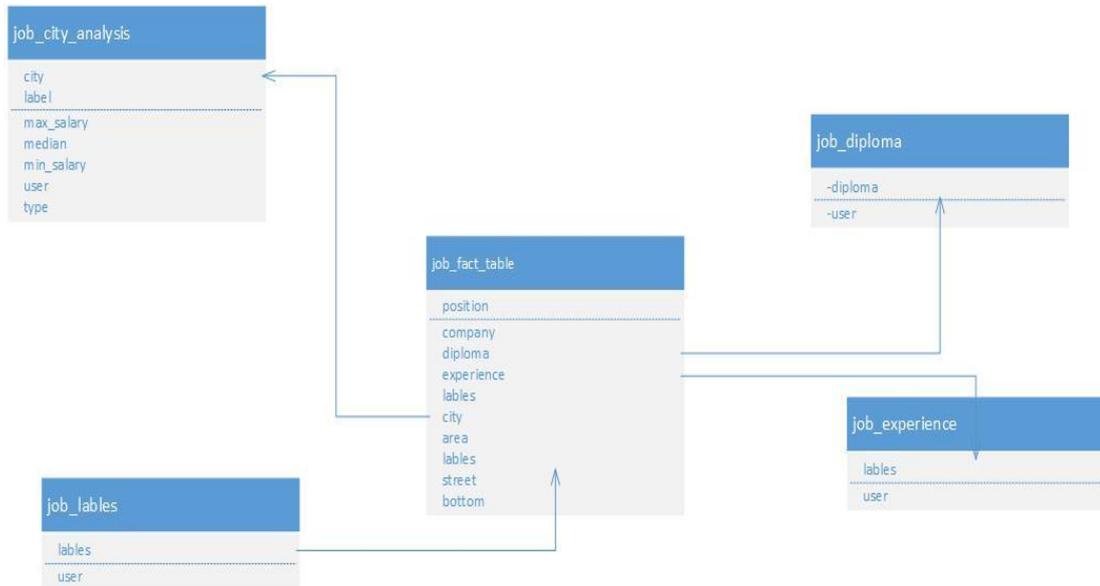


图 4-1 表关系图

4.2 资料表栏位规格书

根据系统功能需求分析,结合系统功能模块设计,对基于大数据技术的招聘网站信息数据分析,每张数据表的名称及相关信息解释如下所示:

- (1) **job_fact_table** 资料表

表 4-1 job_fact_table 资料表栏位规格书

栏位名称	型号	栏位解释
position	text	职位
company	text	公司
salary	text	薪资
diploma	text	学历
experience	text	工作经验
city	text	城市
lables	text	行业标签
area	text	地区
bottom	int(11)	最低工资
top	int(11)	最高工资
avg_salary	int(11)	平均薪资

(2) job_city_analysis 资料表

表 4-2 job_city_analysis 资料表栏位规格书

栏位名称	型号	栏位解释
city	text	城市
label	text	标签
max_salary	decimal(18, 2)	最大薪资
median	int(11)	中位线
min_salary	decimal(18, 2)	最小薪资
user	bigint(20)	用户
type	int(11)	类型

(3) job_city_bottom_top 资料表

表 4-3 job_city_bottom_top 资料表栏位规格书

栏位名称	型号	栏位解释
city	text	城市
bottom	bigint(20)	最低工资
top	bigint(20)	最高工资

(4) job_city_salary 资料表

表 4-4 job_city_salary 资料表栏位规格书

栏位名称	型号	栏位解释
city	text	城市
avg_salary	decimal(18, 2)	平均薪资
median	int(11)	中位数

(5) job_diploma 资料表

表 4-5 job_diploma 资料表栏位规格书

栏位名称	型号	栏位解释
diploma	text	学历
user	bigint(20)	用户

(6) job_diploma_salary_post 资料表

表 4-6 job_diploma_salary_post 资料表栏位规格书

栏位名称	型号	栏位解释
diploma	text	学历
avg_salary	decimal(18, 2)	平均薪资
median	int(11)	中位数
number	bigint(20)	数量

(7) job_salary_experience_post 资料表

表 4-7 job_salary_experience_post 资料表栏位规格书

栏位名称	型号	栏位解释
experience	text	工作经验
avg_salary	decimal(18, 2)	平均薪资
median	int(11)	中位数
number	bigint(20)	数量

5 系统实现

本章节主要介绍了本系统的模块大概介绍和系统页面的展示。

5.1 系统模块介绍

登录：输入账号密码，用户可以登录后浏览招聘网站的信息。

注册：用户注册，注册时用户需要输入账号、密码和确认密码，然后即可注册成功。

数据报表：根据已收集的地区数、公司总数、职位总数以及数据总数，全面分析各地区岗位、公司数量和各地区岗位、公司数量占比。

职位推荐：为了使用户能更加方便快捷地查询和实现简单的基本操作，系统提供分析各个岗位的招聘数据，而且还能根据自身的学历、工作经验以及预期薪资搜寻适合自己的岗位，可以跳转岗位链接迅速了解详情。

地区分析：包含地区综合分析、城市薪资箱位图、平均薪资和岗位数量对比。用户可以浏览关于地区之间对现有的招聘网站信息数据进行的分析处理页面。在地区综合分析中，还可以勾选地区、工作经验、学历、行业进行条件选择查找，用来了解适合的岗位信息，方便用户根据所要的条件进行动态分析，以供参考。

薪资分析：包含薪资与岗位关系、薪资与学历岗位关系和薪资经验岗位关系。用户可以浏览关于薪资方面对现有的招聘网站信息数据进行详细的分析处理页面。

标签分析：包含行业分析、学历占比和工作经验。用户可以浏览关于行业、学历和工作经验的对于现有的招聘网站信息数据进行的分析处理页面。

用户管理：包含用户列表，主要对运用过此系统的用户进行记录，管理者可以增加或删除所有用户的信息。

5.2 系统页面展示

登录：输入账号密码，用户可以登录后浏览招聘网站的信息。

如图 5-1 所示：

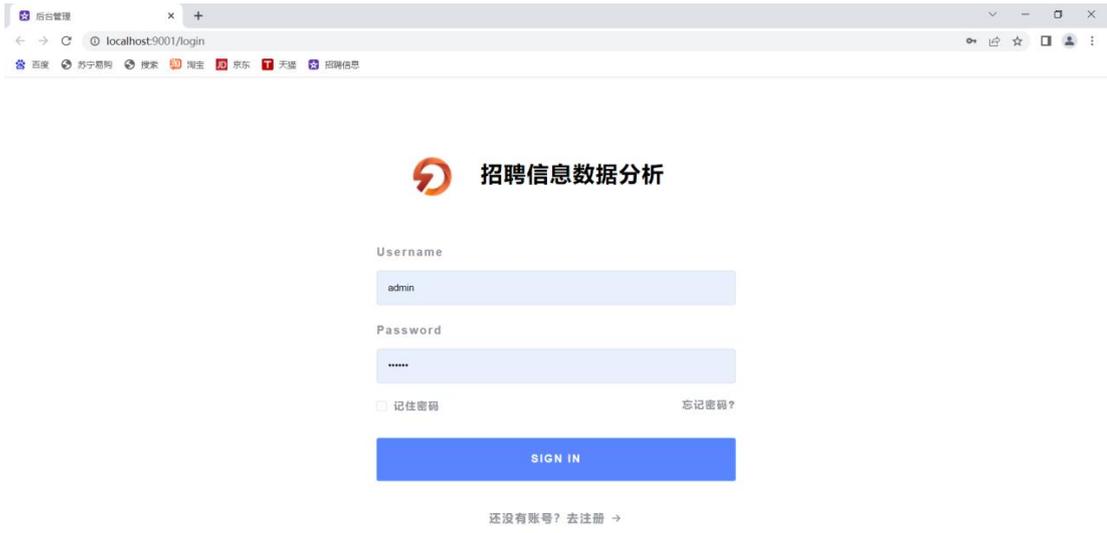


图 5-1 用户登陆界面

注册：用户注册，注册时用户需要输入账号、密码和确认密码，然后即可注册成功。

如图 5-2 所示：

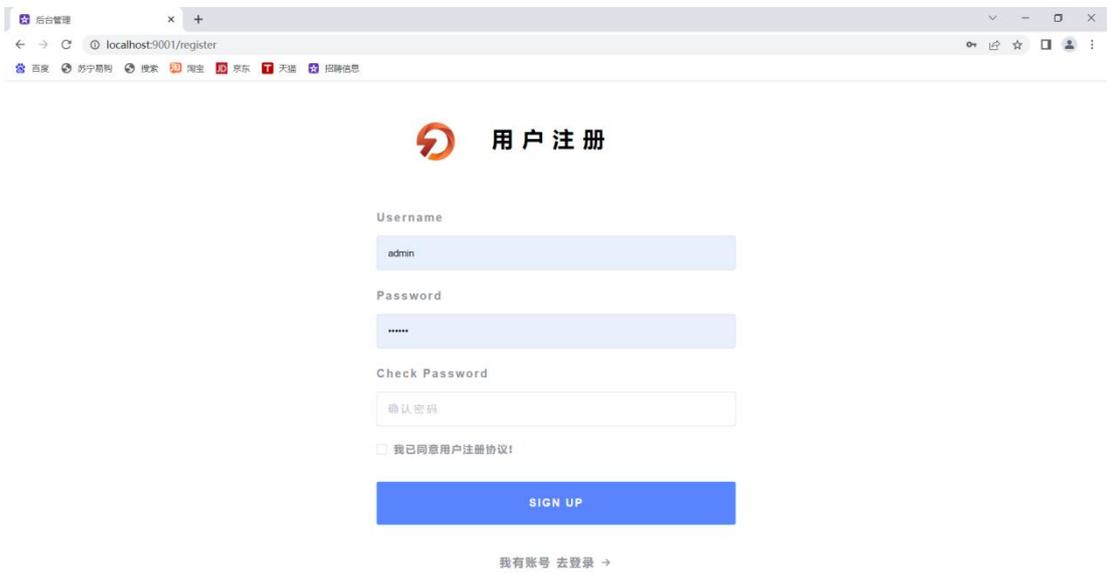


图 5-2 用户注册界面

数据报表：根据已收集的地区数、公司总数、职位总数以及数据总数，全面分析各地区岗位/公司数量和各地区岗位/公司数量占比。

如图 5-3 所示：

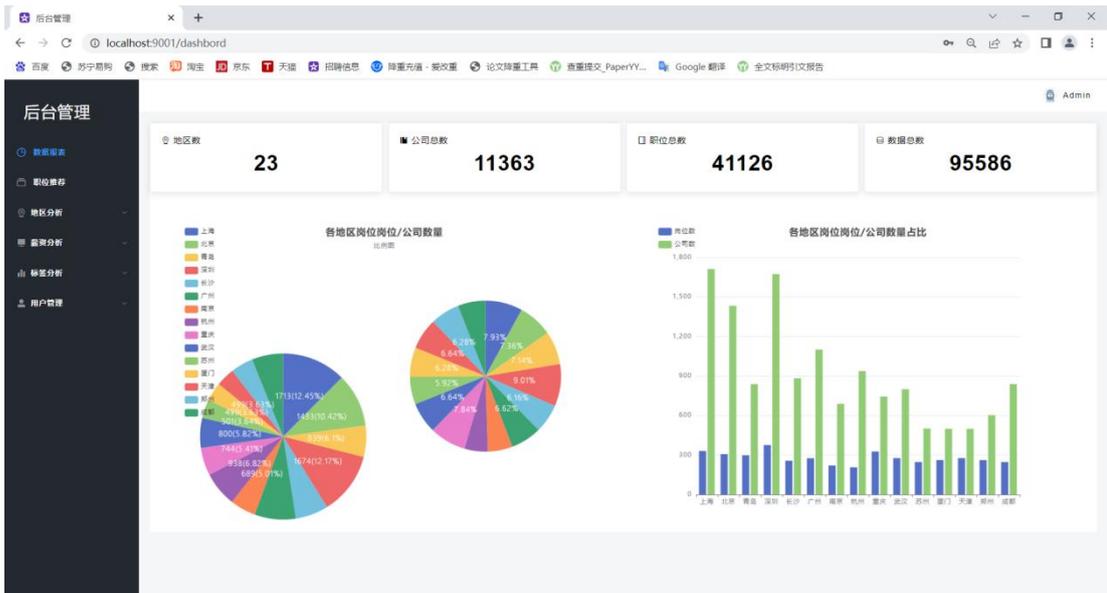


图 5-3 数据报表页面

职位推荐：为了使用户能更加方便快捷地查询和实现简单的基本操作，系统提供分析各个岗位的招聘数据，而且还能根据自身的学历、工作经验以及预期薪资搜寻适合自己的岗位，可以跳转岗位链接迅速了解详情。

如图 5-4 所示：

地区	工作经验	学历	薪资范围(K)	公司规模	福利待遇	链接
杭州	不限	不限	5K-8K	1000-9999人	绩效奖金, 加班补助, 包吃, 弹性工作, 节日福利	了解详情>>
杭州	不限	不限	5K-8K	1000-9999人	绩效奖金, 加班补助, 包吃, 弹性工作, 节日福利	了解详情>>
杭州	不限	不限	8K-10K	20人以下	绩效奖金, 年终分红, 股票期权, 弹性工作, 员工旅游	了解详情>>
杭州	不限	不限	8K-10K	20人以下	绩效奖金, 年终分红, 股票期权, 弹性工作, 员工旅游	了解详情>>
杭州	不限	不限	5K-8K	1000-9999人	每年多次调薪, 带薪年假, 节日福利, 餐补	了解详情>>
杭州	不限	不限	5K-8K	1000-9999人	绩效奖金, 加班补助, 包吃, 弹性工作, 节日福利	了解详情>>
杭州	不限	不限	5K-8K	1000-9999人	绩效奖金, 加班补助, 包吃, 弹性工作, 节日福利	了解详情>>
杭州	不限	不限	5K-8K	1000-9999人	绩效奖金, 加班补助, 包吃, 弹性工作, 节日福利	了解详情>>
杭州	不限	不限	5K-8K	1000-9999人	绩效奖金, 加班补助, 包吃, 弹性工作, 节日福利	了解详情>>
杭州	不限	不限	4K-7K	1000-9999人	绩效奖金, 加班补助, 包吃, 弹性工作, 节日福利	了解详情>>
杭州	不限	不限	4K-6K	1000-9999人	绩效奖金, 加班补助, 包吃, 弹性工作, 节日福利	了解详情>>

图 5-4 职位查找分析页面

地区分析：包含地区综合分析、城市薪资箱位图、平均薪资和岗位数量对比。用户可以浏览关于地区之间对现有的招聘网站信息数据进行的分析处理页面。

在地区综合分析中，可以勾选地区、工作经验、学历、行业进行条件选择查找，用来了解适合的岗位信息，方便用户根据所要的条件进行动态分析。

如图 5-5 所示:



图 5-5 地区分析-地区综合分析查找页面

城市薪资箱位图：该图上可以总览主要城市薪资水平的箱位分布情况。在该图上可以标注和查取某一城市的每一水平薪资的详细数据。

如图 5-6 所示:

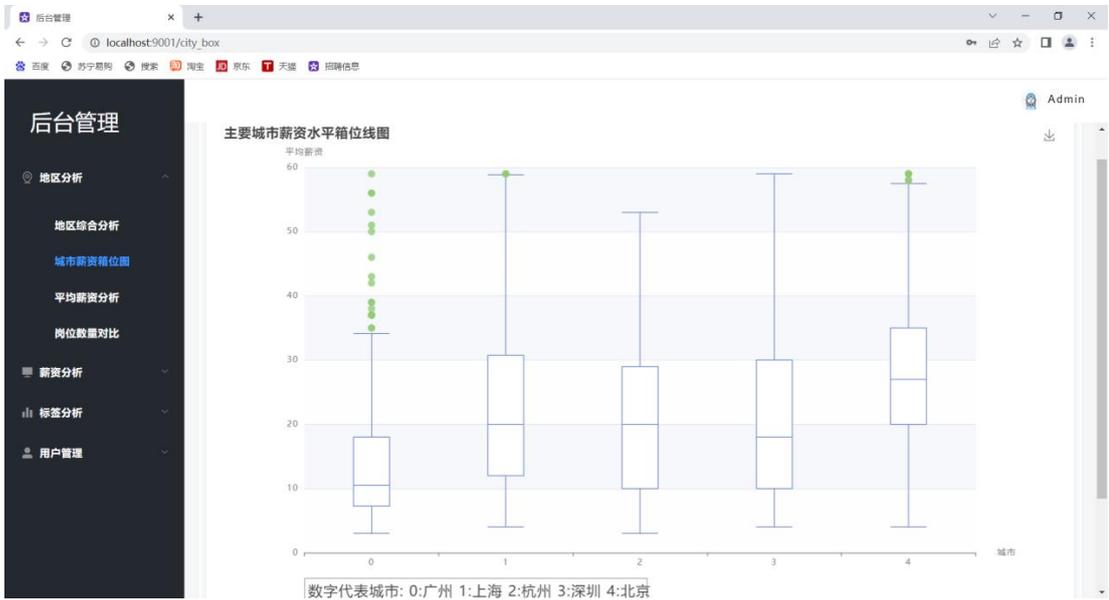


图 5-6 地区分析-城市薪资箱位图

平均薪资分析：对于获取到的热门城市薪资得到中位数和平均数的分析详细数据、对比，可根据自身喜好切换数据表图、折线图和柱状图，以助于更加清晰、直观的了解各地区的薪资水平，比如北京的薪资水平位于第一。

如图 5-7 所示：

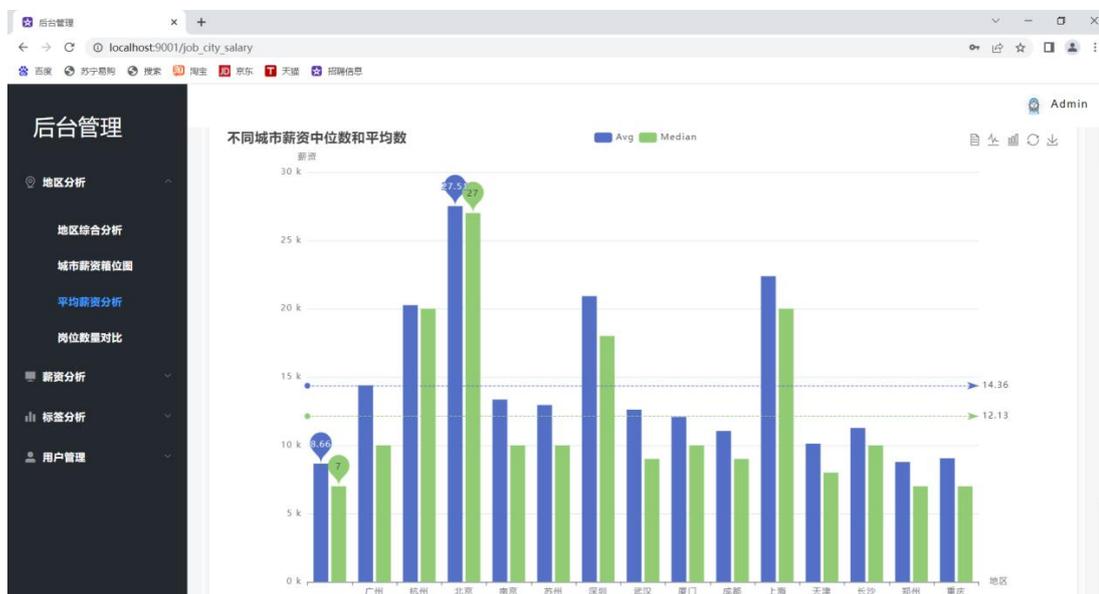


图 5-7 地区分析-平均薪资分析

岗位数量对比：针对各个热门城市，以 10K（10000 人民币）为界限，分析各个地区的岗位数和占比，比如在薪资大于 10K 的岗位数，北京占比最大。

如图 5-8 所示：

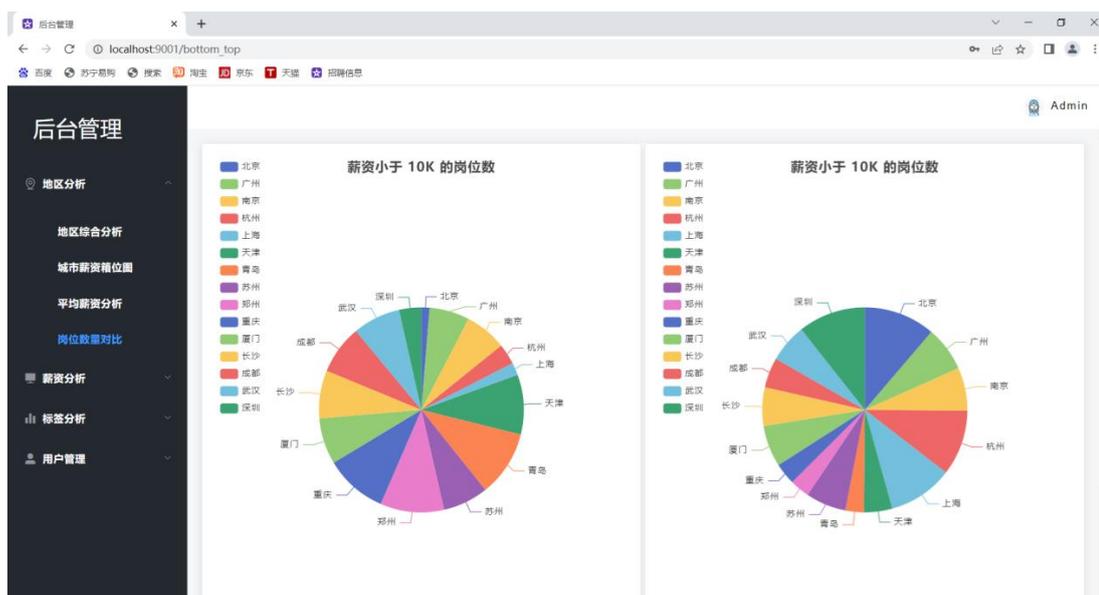


图 5-8 地区分析-岗位数量对比

薪资分析：包含薪资与岗位关系、薪资与学历岗位关系和薪资经验岗位关系。

用户可以浏览关于薪资方面对现有的招聘网站信息数据进行详细的分析处理页面。

薪资与岗位的关系：分析热门城市的招聘信息数据的薪资与岗位数的关系，可以了解到各个薪资阶层的岗位数，比如 5~10K 的岗位数最多。

如图 5-9 所示：

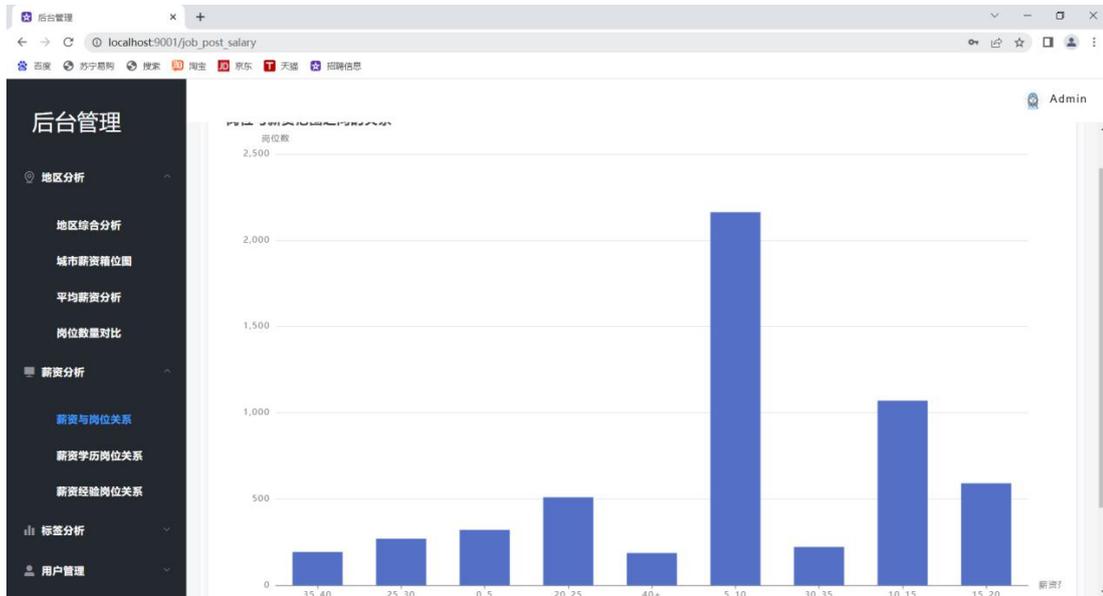


图 5-9 薪资分析-薪资与岗位的关系

薪资学历岗位关系：分析热门城市的招聘信息数据的学历要求与薪资的关系，可以根据自身喜好切换数据表图、折线图和柱状图观察平均数和中位数，了解到各个阶层的学历需求和薪资的关系，比如要求为博士的岗位薪资最高。

根据搜集到的岗位数、学历要求占比和要求的的工作经验来了解到各个地区职位招聘需求现状。

行业分析：用户可以根据收集到的各个行业的岗位数，了解到各个行业的职位招聘需求现状。

如图 5-12 所示：

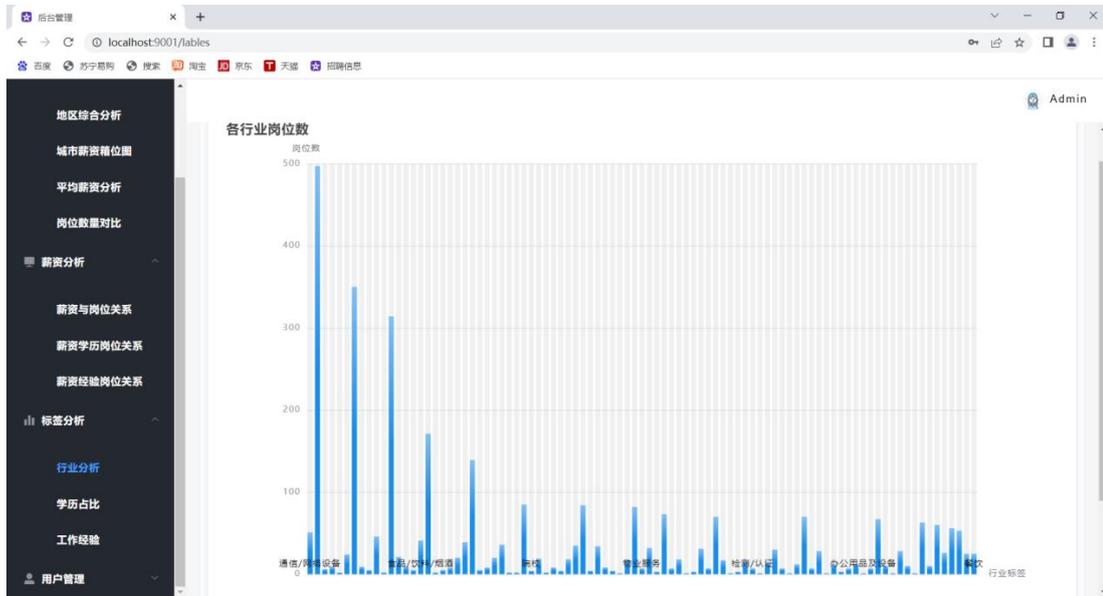


图 5-12 标签分析-行业分析

学历占比：用户可以根据收集到的岗位数中各个阶层的学历要求分析占比情况，比如学历要求为本科的岗位数占比最大。

如图 5-13 所示：

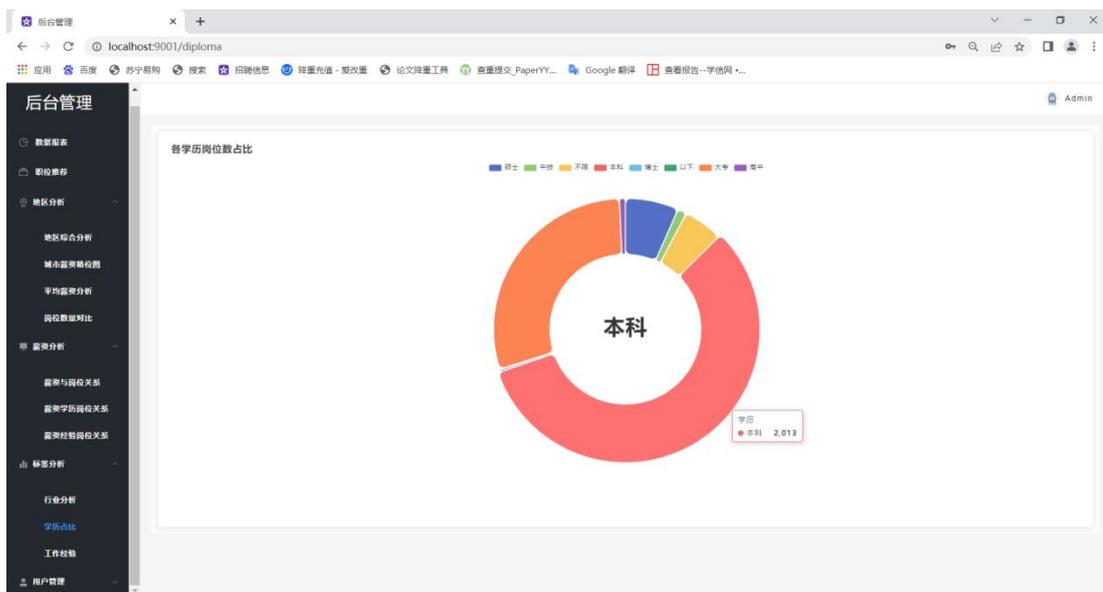


图 5-13 标签分析-学历占比

工作经验：用户可以根据收集到的岗位数中各个阶层的工作经验分析柱状图的情况，比如工作经验为 1~3 年的岗位数最多，工作经验为 10 年以上的岗位数最少。

如图 5-14 所示：

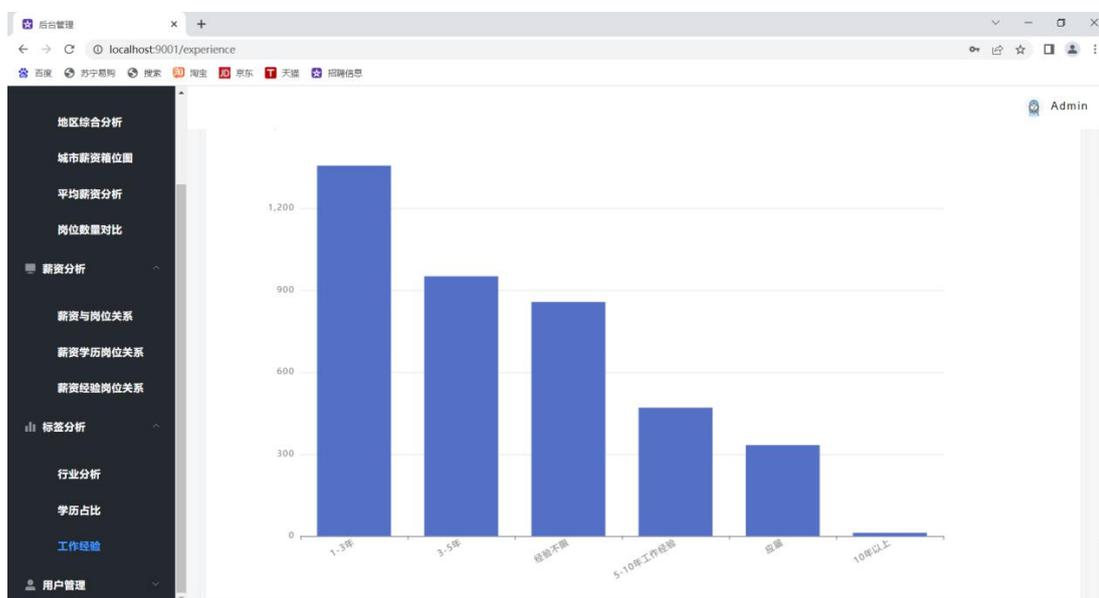


图 5-14 标签分析-工作经验

用户管理：包含用户列表，主要对运用过此系统的用户进行记录，管理者可以增加或删除所有用户的信息。

如图 5-15 所示：

ID	用户名	昵称	年龄	性别	地址
13	zhang	张三	20	女	木星
16	qian	钱江	22	男	地球
17	zhagsan	用户对象			
18	1		3	男	1
25	admin	哈哈哈哈哈			的粉红色的花
26	zhagsan111				
27	as	用户层的			
28	zhagsan4				
29	zhagsan3				

图 5-15 用户管理-用户列表

基本设置：可以根据自身喜好，设置自己账户的昵称、修改用户名和密码、年龄、选择性别和输入地址。如图 5-16 所示：

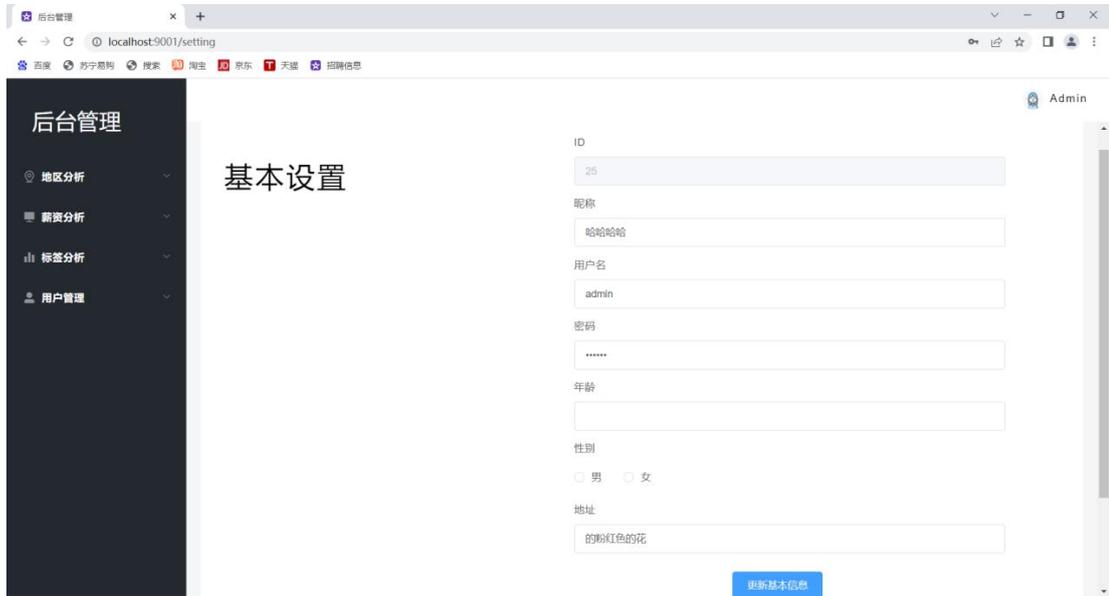


图 5-16 基本设置页面

6 总结及未来展望

6.1 总结

目前，全国最大的招聘网站每天更新的招聘信息达到 3000 多条，上网投递的求职简历达 2 万多份。整个职位信息库保持有效招聘信息达 8 万多近 30 万个空缺职位。最新的研究统计数据表明，全国数十家招聘网站平均每月发布的招聘职位都在数万以上，平均日访问量达到数万人次。随着网络科技的不断发展和网络的广泛应用，运用大数据技术进行招聘信息的分析与应用具有相当大的发展潜力与可行性。

当然，运用传统的方法搜寻招聘信息并不难，但要运用规律的方法收集资料和整合成可视化数据是比较困难的。刚开始，我运用传统的资料收集和原始的人工查询方法，只能看到视觉化现象，无法看到资料背后的格式，并且基本上每个招聘岗位页面的资料格式和招聘需求格式都不一样，所以难点在于找到规律性、一致性的资料格式和如何把收集到的资料直观的可视化展现出来。所以，本系统运用到数据预处理、数据清洗、数据可视化的技术对获取到的数据进行多维度分析，划分学历、薪资等招聘需求进行分组，将半结构化数据处理为结构化数据，运用交错分析和视觉化分析，更直观地了解大量的复杂数据。客观分析我国多个地区岗位的地域分布、热点分布、薪资水平现状，提出研究成果，并展现结果的可视化关系，能够足不出户随时掌握各个招聘信息网站上的最新招聘职位信息数据进行详细的分析。本毕业设计，就是为了适应现代求职、招聘方式而做出的一种创新性尝试，可以有效地解决了招聘工作效率和效果改进提升中所面临的问题。因此，本系统具有低成本、大容量、速度快和强调个性化服务的优势，允许更加灵活的方式，提供更丰富的信息资源，为广大求职者提供便利。

希望此系统能够为各个公司寻找符合其需求的各种专业人才同时也为各个行业、各个地区的人才寻找到一份能够充分体现自身价值的工作。

6.2 未来展望

在老师的帮助下，我顺利完成了最后的项目。我从这个最终项目中学到了很多东西，学会了自己解决问题。但是，由于个人经验和技术问题，该系统也存在不足之处，如岗位数据有待及时更新和提高精确度，代码写得比较繁琐和吃力、有的功能模块还做得不是很全面，比如增加在线咨询服务，即求职者找到感兴趣或适合的岗位后，通过该功能模块想发布招聘信息的公司企业进行在线咨询，可以让求职者进一步了解该公司的基本状况，这些方面有待进一步解决和完善。在以后的工作中，我相信一定会认真考虑到各种需求，未来能够开发更具有智慧性的系统，能够把资料及时性的收集与分析，以后一定会尽最大努力达到产品的理想高度。

参考文献

- [1] 刘畅. 数据类岗位招聘需求信息研究[D]. 兰州财经大学, 2019.
- [2] 岳琦淞. 基于大数据的高校人力资源管理的创新策略[J]. 科技与创新, 2021(5):125-127, 130.
- [3] 牛海燕. 前程无忧校园招聘服务改善研究[D]. 黑龙江:哈尔滨工业大学, 2019.
- [4] 令霞. 简议大数据与信息管理[J]. 传播力研究, 2018(24):1.
- [5] 平淑容. 云存储下高效安全的可搜索加密方案研究[D]. 江苏:江苏大学, 2018.
- [6] 王涛. 基于 Python 的软件技术人才招聘信息分析与实现——以前程无忧为例[J]. 福建电脑, 2018, 34(11):118-119.
- [7] 王晶. 关于舆情监测系统实时交互式可视化的研究[D]. 陕西:西安理工大学, 2019.
- [8] 陈美君, 胡小猛. 现代旅游信息管理系统(TIMs)的构建[J]. 旅游科学, 2004, 18(4):5.
- [9] 林华. 基于 WEB 的教务管理系统的安全性设计与实现[D]. 重庆大学, 2005.
- [10] 伍冠军. 公安情报研判分析平台的设计与实现[D]. 四川:电子科技大学, 2010.
- [11] 刘学军, 李长云, 万烂军. 基于 Spark 的工业大数据处理可视化平台应用研究[J]. 福建电脑, 2017, 33(12):3.
- [12] 顾紫冉, 廖文焘. 基于 RFID 的智能图书馆图书识别管理与分类系统的设计[J]. 电子元件与信息技术, 2020, 13(4):4.
- [13] 彭赓, 范明钰. 基于改进网络爬虫技术的 SQL 注入漏洞检测[J]. 计算机应用研究, 2010(7):3.
- [14] 孙建立, 贾卓生. 基于 Python 网络爬虫的实现及内容分析研究[C]// 中国计算机用户协会网络应用分会 2017 年第二十一届网络新技术与应用年会论文集. 2017.
- [15] 孙舵. 基于 ASP.Net 的网上招聘求职系统设计与实现[D]. 吉林大学.
- [16] 原舒雨, 蔡皖东, 李勇军, 等. 面向博客的智能网络蜘蛛技术及其系统实现[J]. 微电子学与计算机, 2011, 28(1):4.
- [17] 谢克武. 大数据环境下基于 python 的网络爬虫技术[J]. 电子制作, 2017(9):2.
- [18] 李昌. Web 应用安全防护技术研究及实现[D]. 中南大学, 2010.
- [19] 薛滨. 刍议大数据背景下企业财务管理的挑战与变革[J]. 商场现代化, 2019(4):2.
- [20] 宋学清, 刘雨. 大数据:信息技术与信息管理的变革[J]. 情报科学, 2014.

致谢

经过两个多月的努力，我从最开始的茫然，到渐渐的进入状态，再到具体思路清晰明了，整个论文写作和毕设创作阶段，都很难用语言表达。紧张而又有序的毕业论文写作和毕业设计创作终于落下帷幕。

回顾这段时间，我思绪万千，从毕业论文（设计）选题再到开题，我认真上网查找资料，仔细筛选文章文献，为随后的论文写作做好准备。从11月开始，我便试着罗列出提纲，并和曹永忠老师沟通交流确定好题目是：基于大数据技术的招聘网站信息数据之分析与应用。后来，我开始根据所收集到的材料开始进行开题报告的撰写，并与导师召开开题报告会，并进一步修改完善开题报告和任务书。在导师耐心指导下，我对论文的写作方向有了进一步的认识。在查找资料的过程中，遇到困难我及时和指导老师沟通，与同学们互相交流。在大家的帮助下和自身的辛苦努力下，问题解决，毕业设计越来越完善，论文也慢慢的成型。在整个过程中，我学习到了很多很多的知识，专业水平大大提升，增长了见识。

我清楚地感觉到，我们的生活离不开网络技术的不断发展和网络的广泛应用。互联网以其独特的优势征服了我们。进入21世纪，网络不断发展。随着Web技术的不断进步，使得任务变得非常琐碎且容易出错。在这种情况下，需要一种切实可行的方法来分析和利用招聘网站的信息数据。

此系统根据大数据理论，以部分热门城市的相关招聘信息为数据源，分析中国多个岗位的地区分布、工作经验要求、学历要求、薪资水平现状。

在信息时代，要创建和传播的数据量将变得越来越庞大，而这些数据将构成大量社会财富的基础。因此，在大数据时代背景下，计算机信息处理能力需要不断提高和提高，计算机网络工作人员应该利用这个机会。加强各类数据信息处理技术的精细化和创新，迎接大数据时代的新挑战。

在整个学习期间，我能够更好地了解如何使用pycharm2020.3和idea2020.3，学习编程方法和技巧。技术上，我学到了很多新东西，尤其是JavaWeb、MySQL 5.7、pycharm 2020.3。我意识到系统开发需要综合考虑方方面面，而不仅仅是写代码。同时，认真对待每件事都需要自己的综合专业知识。最后，感谢老师们抽出宝贵的时间审阅这篇论文。以后我会更加努力学习，不辜负大家的期待！

附录

附录 1

闽江学院

本科毕业论文（设计）文献综述

应用
题目 基于大数据技术的招聘网站信息数据之分析与

学院 数学与数据科学学院（软件学院）

年 级 2018 级

专 业 软件工程

学生姓名 陈晞玟

学 号 3187103111

指导教师 曹永忠

职 称 教授

基于大数据技术的招聘网站信息数据之分析与应用问题研究

文献综述

1 前言

本人毕业设计的题目为《基于大数据技术的招聘网站信息数据之分析与应用》，这些文献给与本文很大的参考价值。本文主要查阅近几年有关招聘网站信息数据的文献期刊。

大数据是近年来出现在通信和计算机领域中的一个热门关键词。关于大数据，尚未有一个统一的定义，但却有两个观点能够诠释大数据的本质。第一个观点来自于 Gartner 公司的 Merv Adrian 在 2011 年第一季度刊登在 Teradata Magazine 上的一篇文章，文中指出“数据超出了常用硬件环境和软件工具在可接受的时间内为其用户收集、管理和处理数据的能力” [1]。另一个观点来自于麦肯锡全球数据分析研究所 (Mckinsey Global Institute) 在 2011 年 6 月发布的《大数据：创新、竞争和生产力的下一个前沿》报告，报告中提出“大数据是指大小超出了典型数据库软件工具收集、存储、管理和分析能力的数据集” [2]。麦肯锡称：“数据，已经渗透到当今每一个行业和业务职能领域，成为重要的生产因素。人们对于海量数据的挖掘和应用，预示着新一波生产率增长和消费者盈余浪潮的到来。”

岳琦淞 [9] 提到大数据已经深刻地影响到人们的生活、工作和学习。大数据的意义在于对由多种类型数据构成的数据集体进行分析和研究，提取有利用价值的信息，从而帮助人们在解决问题时可以作出科学的决策。同样大数据的威力强烈地冲击着招聘系统，正在成为推动招聘系统创新与变革的颠覆性力量。

随着信息技术的飞速发展和日趋激烈的商业竞争，人力资源管理的处理也逐渐采用科学化的管理。[10] 本论文论述了人力资源管理系统开发过程：需求分析、功能设计、数据库设计、系统实施。对系统结构、功能设计的流程做了全面而详尽的阐述，然后针对系统的各功能模块做了具体的分析与研究。通过对功能及数据流的分析，得到系统数据库的逻辑结构，并在此基础上完成数据库物理结构的创建，从而实现对各种信息的查询等作用。[12] 整体提高了管理效率和对信息的可控性，降低管理成本，提高执行力，使管理趋于完善。

2 现状研究

2.1 大数据的意义

大数据,顾名思义,指的是数据的大量采集,它需要在采集和记录大量数据的基础上才能进行数据库的搜索匹配。一般而言,我们认为大数据的内涵包括三个方面:(1)数据量巨大、来源广泛、类型多样;(2)需要用时下最先进的技术进行数据分析;(3)通过数据分析挖掘出匹配的价值[29]。

随着云物移大智等新一代信息技术的发展普及,政务大数据,商业大数据,医疗大数据,教育大数据,科学大数据等多领域的的数据资源不断丰富和持续累积,各国都在加快大数据战略布局,抢占大数据制高点[28]。

王震一[8]提出:今天的大数据就像当年发明显微镜一样,人们从庞杂的海量数据中找到了前所未知的事物。正确面对这些关系复杂、形式多样的结构化、半结构化和非结构化的教育数据,形成一套涵盖业务、技术和IT基础架构的全面解决方案来处理存储、管理和分析教育大数据,这就是信息化教育。

2.2 数据类岗位需求的重要性

正如刘畅[3]所说:随着高等教育的发展与普及,我国高校毕业生数量逐年增加,屡创历史新高;在大数据与人工智能的发展进程中,数据类岗位的供需缺口也日渐加大。但是,由于人才招聘中存在结构性招聘矛盾和招聘信息不对称的问题,常出现毕业生难以找到理想工作,企业难以招聘到理想人才的困境。在此背景下,互联网招聘打破了地理空间的限制,扩大了供需双方的范围,成为招聘者发布信息与应聘者获得信息的重要渠道。得益于网络招聘的广泛应用,互联网招聘信息的实时发布使人才招聘中的信息不对称问题得到改善,但招聘文本仍具有海量,非结构化的特点,在文本统计分析中具有一定的技术困难。提取网络招聘文本的主题词,分析大数据时代数据类岗位的招聘需求,对广大数据科学相关专业的高校毕业生更高质量和更充分的就业有积极的促进作用。基于此,本文利用互联网招聘数据对数据类岗位招聘需求信息进行系统研究,文章首先选取互联网招聘平台中全国范围内的数据类岗位招聘数据,运用自然语言处理技术对非结构化的招聘信息进行文本预处理;进而结合LDA主题模型与word2vec词嵌入技术构建了招聘需求主题词提取模型,将招聘需求分解为教育背景,工作经验,知识技能与个人素质四个主题,提取出各数据类岗位基于不同需求主题的招聘需求主题词。

2.3 招聘服务需求

随着移动互联网的飞速发展,企业之间的竞争正在从资金竞争转向人才竞争。为了占据人才优势,保持公司长久的驱动力及竞争力,企业意识到人才储备对未来发展至关重要。校园招聘作为人力资源招聘模式之一,已经成为了企业人力资

源管理中最核心的业务之一[27]。随着校园招聘业务需求的增加,提供此项服务的专业人力资源服务商随之增多,如前程无忧(51Job)、智联招聘(Zhaopin)、中华英才网(ChinaHR)、猎聘网(Liepin)等等。前程无忧作为国内领先的专业人力资源服务商,在面对国内就业难度及同行业竞争激烈的局势,应该提供何种形式的校园招聘服务,如何改善目前校园招聘服务里存在的问题,是目前前程无忧公司在校园招聘业务发展方向上考虑的重点问题[4]。希望通过对校园招聘服务方面的改善,提高企业校园招聘有效性,提高客户满意度,进而提高前程无忧校园招聘部门的合作业绩,在全国校园招聘服务市场占有更多的份额。

3 技术应用研究

3.1 Java web 的运用

近些年,互联网技术飞速发展,越来越多的人接触到了编程语言,同时更多的人愿意去了解学习编程语言[12]。

杨小杰[7]提出在当前的主流环境下,电子商务已经朝着多样化、多元化的方向发展,人们在网络上购买商品的优势正在逐渐凸显。网上购物系统作为新型的商业模式越来越彰显出其新兴的活力,极大的促进了传统商业模式向互联网经营模式的靠拢。分析了基于WEB的信息在线编辑器的使用、升级及集成方法。分析了一个编辑器需要整合的多种技术[13]。从图形图像、声音、视频流媒体、公式编辑四个技术方向着重介绍了目前所使用的流行技术,及相应的在线编辑技术。

随着网络的迅猛发展和高校的扩招,高校的管理方法和教学模式均发生了很大的变化。信息技术的飞速发展,为人类的发展提供了快速、准确、可靠的信息交流渠道,改变人们的工作方式和固有的思维习惯[14]。

3.2 Spring 的运用

刘彬[6]提到,随着互联网技术的飞速发展,计算机及其相关技术在社会各个方面的应用也越来越广泛,其发挥的巨大作用是也是不可估量和替代的。

胡世港[5]提出在管理业务领域,整理出详细的软件功能需求,形成软件需求分析报告.在需求分析的基础上,建立了管理系统的领域模型,并决定采用SpringMVC作为底层开发框架.同时,对该系统中的业务流程管理提出了引入activiti 工作流框架的展望。

4 总结

此次,可以认为从理论层面探讨大数据下的招聘信息数据分析与运用这一课题,必须要有理论的支持,同时可以试图定义两者之间的关系,这对后来者进行

系统的研究具有重要的意义[16]。

现在借助大数据进行招聘已经是高等学校迫切需要解决的课题，如何随时掌握各个招聘信息网站上的最新招聘职位信息数据的进行详细的分析做这样的课题研究显得尤为有意义。我们以全面的眼光研究这个课题，开发适合各个时代背景下大数据的招聘工作的研究模型，同时鼓励各个企业在这个系统的基础上进行精加工以更贴合本学科的学科特色成为亟待解决的问题。

参考文献

- [1] Merv Adrian. Big Data:it' s going mainstream and it' s your next opportunity[J]. Teradata Magazine, 2011,(1):3-5.
- [2] Manyika J, Chui M, Brown B, et al. Big data: The next frontier for innovation, competition, and productivity[R].USA:Mckinsey Global Institute, 2011.
- [3] 刘畅. 数据类岗位招聘需求信息研究[D]. 兰州财经大学, 2019.
- [4] 牛海燕. 前程无忧校园招聘服务改善研究[D]. 黑龙江:哈尔滨工业大学, 2019.
- [5] 胡世港. 基于 SpringMVC 框架的电子档案管理系统的设计与实现[J]. 电脑知识与技术:学术版, 2018(9X):2.
- [6] 刘彬. 基于 Spring MVC 框架的茶楼管理系统的设计与实现[D]. 吉林大学.
- [7] 杨小杰. 基于 JavaEE 的网上购物系统的设计与实现[D]. 吉林大学.
- [8] 王震一. 教育离“信息化”到底还有多远[J]. 中小学信息技术教育, 2012, (12):25-26.
- [9] 岳琦淞. 基于大数据的高校人力资源管理的创新策略[J]. 科技与创新, 2021(5):125-127, 130.
- [10] 令霞. 简议大数据与信息管理[J]. 传播力研究, 2018(24):1.
- [11] 平淑容. 云存储下高效安全的可搜索加密方案研究[D]. 江苏:江苏大学, 2018.
- [12] 王涛. 基于 Pyhton 的软件技术人才招聘信息分析与实现——以前程无忧为例[J]. 福建电脑, 2018, 34(11):118-119.
- [13] 王晶. 关于舆情监测系统实时交互式可视化的研究[D]. 陕西:西安理工大学, 2019.
- [14] 陈美君, 胡小猛. 现代旅游信息管理系统(TIMS)的构建[J]. 旅游科学, 2004, 18(4):5.
- [15] 林华. 基于 WEB 的教务管理系统的安全性设计与实现[D]. 重庆大学, 2005.
- [16] 伍冠军. 公安情报研判分析平台的设计与实现[D]. 四川:电子科技大学, 2010.
- [17] 刘学军, 李长云, 万烂军. 基于 Spark 的工业大数据处理可视化平台应用研究[J]. 福建电脑, 2017, 33(12):3.
- [18] 顾紫冉, 廖文焘. 基于 RFID 的智能图书馆图书识别管理与分类系统的设计[J]. 电子元件与信息技术, 2020, 13(4):4.
- [19] 彭赓, 范明钰. 基于改进网络爬虫技术的 SQL 注入漏洞检测[J]. 计算机应用研究, 2010(7):3.
- [20] 孙建立, 贾卓生. 基于 Python 网络爬虫的实现及内容分析研究[C]// 中国计算机用户协会网络应用分会 2017 年第二十一届网络新技术与应用年会论文集. 2017.
- [21] 孙舵. 基于 ASP.Net 的网上招聘求职系统设计与实现[D]. 吉林大学.
- [22] 原舒雨, 蔡皖东, 李勇军, 等. 面向博客的智能网络蜘蛛技术及其系统实现[J]. 微电子学与计算机, 2011, 28(1):4.
- [23] 谢克武. 大数据环境下基于 python 的网络爬虫技术[J]. 电子制作, 2017(9):2.
- [24] 李昌. Web 应用安全防护技术研究及实现[D]. 中南大学, 2010.
- [25] 薛滨. 刍议大数据背景下企业财务管理的挑战与变革[J]. 商场现代化, 2019(4):2.
- [26] 宋学清, 刘雨. 大数据:信息技术与信息管理的又一次变革[J]. 情报科学, 2014.
- [27] 刘哲. 大数据时代下网络招聘有效性研究[D]. 湖南:中南林业科技大学, 2017.

DOI:10.7666/d.Y3269425.

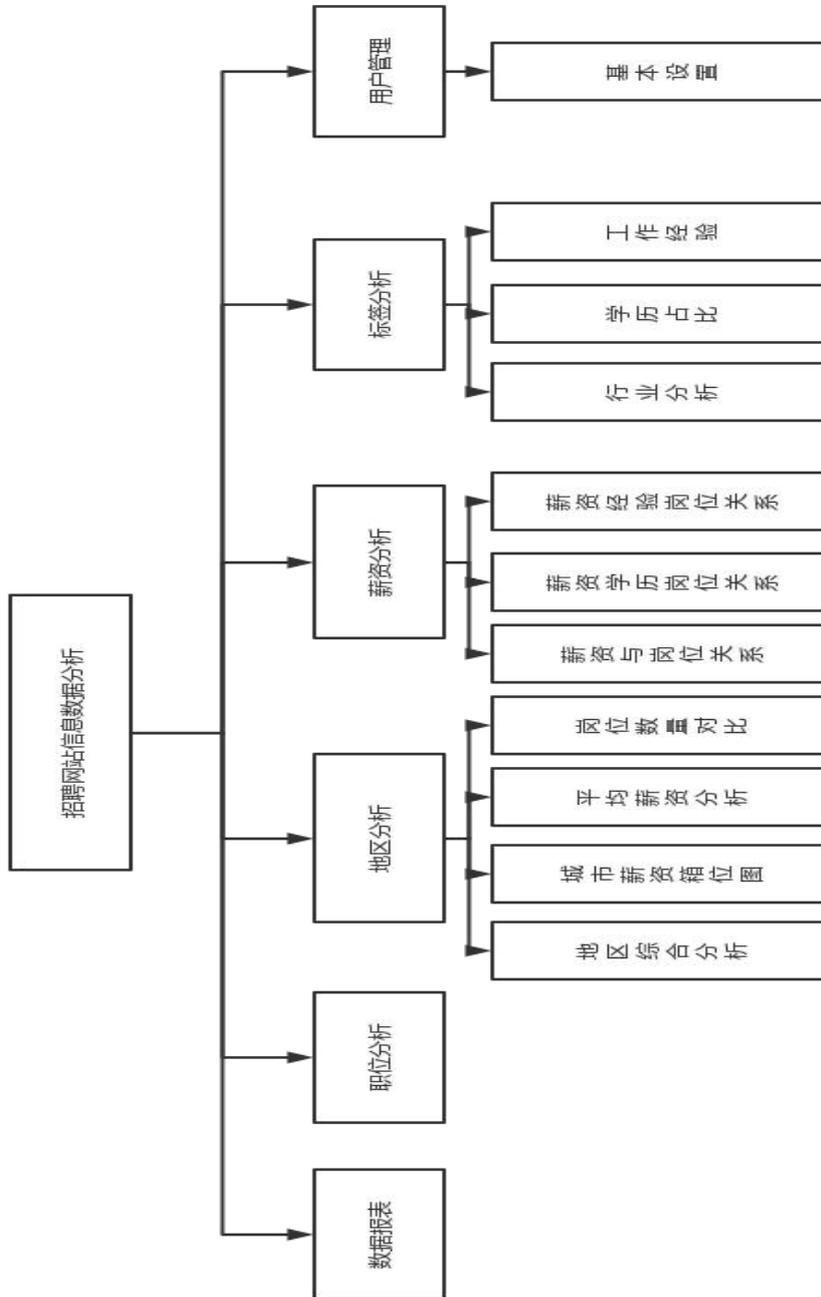
[28] 常绍来. 大数据时代 A 互联网公司人力资源招聘与绩效考核管理研究[D]. 北京:北京化工大学, 2019.

[29] 郑娜. 大数据时代下企业网络招聘的有效性研究——以 C 公司为例[D]. 北京:对外经济贸易大学, 2021.

[30] 陈佳楠. 招聘网站中数据分析类岗位的现状及其影响因素[D]. 广西:广西师范大学, 2020.

附录 2

系统架构图：



附录 3

资料表 SQL 叙述:

表附录 3-1 job_fact_table 资料表 SQL 叙述

```
CREATE TABLE `job_fact_table` (  
  `position` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,  
  `company` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,  
  `salary` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,  
  `diploma` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,  
  `experience` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,  
  `lables` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,  
  `city` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,  
  `area` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,  
  `street` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,  
  `bottom` int(11) NULL DEFAULT NULL,  
  `top` int(11) NULL DEFAULT NULL,  
  `month` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,  
  `commision_pct` decimal(18, 2) NULL DEFAULT NULL,  
  `avg_salary` int(11) NULL DEFAULT NULL  
) ENGINE = InnoDB CHARACTER SET = utf8 COLLATE = utf8_general_ci ROW_FORMAT  
= Dynamic; PRIMARY KEY (`id`),  
  KEY `sysdatetime` (`sysdatetime`)  
) ENGINE=InnoDB DEFAULT CHARSET=latin1 COMMENT='吳厝國小樹屋風速資訊表'  
AUTO_INCREMENT=1;  
--
```

表附录 3-2 job_city_analysis 资料表 SQL 叙述

```
CREATE TABLE `job_city_analysis` (  
  `city` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,  
  `label` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,  
  `max_salary` decimal(18, 2) NULL DEFAULT NULL,  
  `median` int(11) NULL DEFAULT NULL,  
  `min_salary` decimal(18, 2) NULL DEFAULT NULL,  
  `user` bigint(20) NOT NULL,  
  `type` int(11) NOT NULL  
) ENGINE = InnoDB CHARACTER SET = utf8 COLLATE = utf8_general_ci ROW_FORMAT = Dynamic;--
```

表附录 3-3 job_city_bottom_top 资料表 SQL 叙述

```
CREATE TABLE `job_city_bottom_top` (  
  `city` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,  
  `bottom` bigint(20) NULL DEFAULT NULL,  
  `top` bigint(20) NULL DEFAULT NULL
```

```
) ENGINE = InnoDB CHARACTER SET = utf8 COLLATE = utf8_general_ci ROW_FORMAT = Dynamic;
```

表附录 3-4 job_city_salary 资料表 SQL 叙述

```
CREATE TABLE `job_city_salary` (  
  `city` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,  
  `avg_salary` decimal(18, 2) NULL DEFAULT NULL,  
  `median` int(11) NULL DEFAULT NULL  
) ENGINE = InnoDB CHARACTER SET = utf8 COLLATE = utf8_general_ci ROW_FORMAT = Dynamic;
```

表附录 3-5 job_diploma 资料表 SQL 叙述

```
DROP TABLE IF EXISTS `job_diploma`;  
CREATE TABLE `job_diploma` (  
  `diploma` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,  
  `user` bigint(20) NOT NULL  
) ENGINE = InnoDB CHARACTER SET = utf8 COLLATE = utf8_general_ci ROW_FORMAT = Dynamic;
```

表附录 3-6 job_diploma_salary_post 资料表 SQL 叙述

```
CREATE TABLE `job_diploma_salary_post` (  
  `diploma` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,  
  `avg_salary` decimal(18, 2) NULL DEFAULT NULL,  
  `median` int(11) NULL DEFAULT NULL,  
  `number` bigint(20) NOT NULL  
) ENGINE = InnoDB CHARACTER SET = utf8 COLLATE = utf8_general_ci ROW_FORMAT = Dynamic;
```

表附录 3-7 job_salary_experience_post 资料表 SQL 叙述

```
CREATE TABLE `job_salary_experience_post` (  
  `experience` text CHARACTER SET utf8 COLLATE utf8_general_ci NULL,  
  `avg_salary` decimal(18, 2) NULL DEFAULT NULL,  
  `median` int(11) NULL DEFAULT NULL,  
  `number` bigint(20) NOT NULL  
) ENGINE = InnoDB CHARACTER SET = utf8 COLLATE = utf8_general_ci ROW_FORMAT = Dynamic;
```