

基于大数据的就业信息智能分析平台的设计与研究

毛红霞 鲍正德 李晨曦

四川大学锦城学院计算机与软件学院 四川 成都 611731

【摘要】 高校毕业生的就业工作是高校的一项重要工作，因此构建一个就业信息智能分析平台对于高校就业工作来说非常重要。本文将大数据技术运用到大数据就业信息智能分析平台中，对该系统进行了总体设计，并介绍了该系统的主要功能和核心技术。

【关键词】 就业信息；大数据；高校；智能分析

引言

随着 2017 年高校毕业生人数再创新高，就业压力持续不减。然而，高校毕业生对自身职业规划把握不够，在互联网海量的就业信息当中大浪淘沙，难以匹配到符合自身实际情况的工作。

高校毕业生能否对口、合理就业已经成为社会发展与稳定的一件大事，是衡量高等院校办学水平的关键指标^[1]。毕业生的就业质量将直接关系到高校的生存、发展和社会声誉，关系到整个社会的稳定^[2]。因此，高校就业工作是高等院校的一项重要工作，就业信息管理平台是学校就业指导、管理和服务的重要手段。就业信息管理平台应用效果的关键是招聘职位的数量和质量，其核心是能根据学生的就业期望，实现职位的有效推荐，因此利用软件工程的思想 and 统一建模工具，开发一套基于大数据的就业信息分析系统显得尤为重要。

本研究着手利用大数据技术在就业信息智能分析平台中的应用，并创新性地利用网络爬虫技术获取数据，为本平台的数据研究提供参考。在基于前期对平台需求分析和总体设计的基础上，运用大数据爬虫技术、大数据架构技术及数据库技术等对大数据就业信息智能平台进行了总体设计和功能实现，全方位获取企业对学生的招聘信息，汇总分析招聘信息，把握学生用户的就业属性，提升学生检索招聘信息的效率。

1 基于大数据的就业信息智能分析平台的总体设计

为了降低开发和维护的成本、便于各权限用户能够随时随地访问，该就业信息分析平台采用三层 B/S 模式来设计，分为如下三层，如图 1 所示。

1.1 数据存储层

数据存储层用于存放该系统所涉及到的各项数据，为该提供的各项应用提供数据支持。包括：毕业生基本

信息、用人单位招聘信息、高校就业服务系统用户信息。

1.2 服务处理层

服务处理层为用户提供智能化和个性化的服务。通过接收用户在该系统中访问过程中的各种数据操作请求，服务处理层通过与数据存储层的连接，执行相应的 SQL 语句操作，将返回的数据结果进行分析、加工反馈给用户^[3]。

1.3 用户交互层

用户不需要额外安装软件，只需通过浏览器，用户在地址栏输入服务器网址就可以访问系统。用户交互层根据用户的不同权限提供不同的服务接口，包括：用户登录、个人信息管理、求职信息检索、招聘信息汇总分析等功能^[4]。

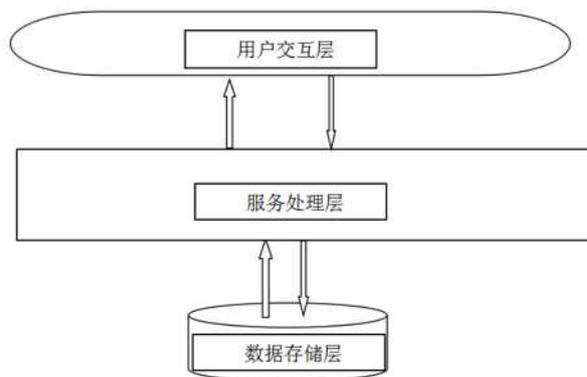


图 1 系统总体结构设计图

2 基于大数据的就业信息智能分析平台的总体功能及核心实现技术

基于大数据的就业信息智能分析平台的主要核心功能，包括数据采集，数据清洗，数据分析，数据展示。

(1) 数据采集：主要用于获取多种数据来源：

用户输入数据和网络数据。针对用户输入的数据使用 MySQL 数据库进行存储；网络数据通过网络爬虫进行网络数据的采集。本系统运用 scrapy-redis 分布式爬虫框架同时将 58 同城、智联招聘、前程无忧等网站校园招聘数据爬取。爬虫结构如图 2 所示：

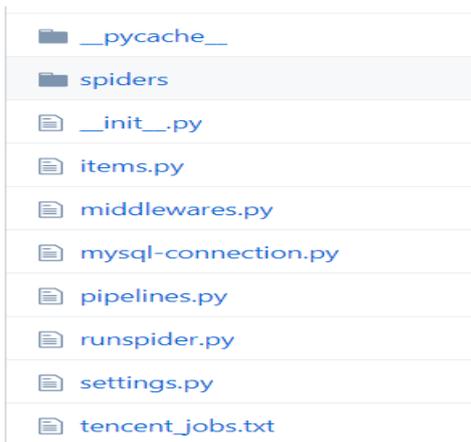


图 2 网络爬虫结构图

(2) 数据清洗：该功能针对从网络上获取的数据进行数据清洗，保证数据的有效性和完整性，并将清洗后的数据存储在数据库中。

网络数据需要经过清洗之后才能保存，用于后续的数据分析。本系统针对网络数据清洗和过滤规则如下：

- ①过滤掉缺少关键数据的字段值。
- ②过滤掉无效数据。
- ③去掉异常值。

下图 3 为清洗公司类型为空的代码截图。

```
#把公司类型为空的数据处理掉
def Data_three(total_one):
    total_two=[]
    for i in total_one:
        if i['公司类型']==None:
            continue
        else:
            total_two.append(i)
    data=pd.DataFrame(list(total_two))
    return data
```

图 3 清洗公司类型为空的代码截图

(3) 数据分析：针对数据存储层中的相关数据进行多维度进行分析，如下：

- ①统计各岗位的招聘信息数量；

- ②统计各岗位的招聘信息地理位置；
- ③统计各岗位的招聘信息的薪资水平等；
- ④统计各职位所需掌握的前 10 位技能点；

(4) 数据展示：从数据分析后端接口获取数据后，使用 Apache Zeppelin 对数据进行可视化展示。

用户通过可视化界面，交互式的输入命令、代码提交给 Zeppelin 编译执行。zeppelin 结合 hadoop/python/spark 处理数据后的进行图表展示返回给用户。Zeppelin 提供可视化机制，通过各类图表展示数据，十分方便，如图 4。通过 zeppelin 的 notebook 一键分享功能，提供指定的 http 给外部进行访问。

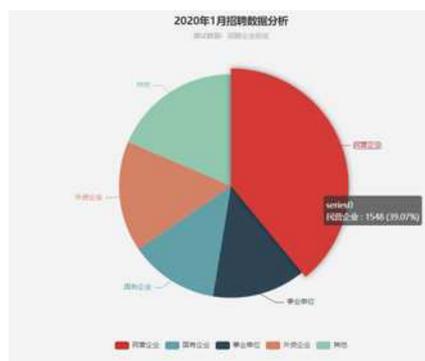


图 4 部分数据展示效果图

3 结束语

基于互联网大数据就业信息智能分析平台，囊括了从数据采集、数据清洗、数据分析到数据展示的所有环节，本文描述了该平台的总体设计以及平台能实现的相关功能，并描写了各功能实现的核心技术。通过该平台的使用，学生能过检索相关岗位的各项数据以及岗位需求，为学生精准就业提供了保障。

【参考文献】

[1] 黄惠萍. 罗定职业技术学院就业信息平台的设计与实现 [D]. 成都: 电子科技大学, 2015:43-44.
 [2] 梁利亭. “互联网+”时代高校毕业生“智慧就业”服务平台研究 [J]. 常州信息职业技术学院学报, 2019, 18(1):23-26+30.
 [3] 董燕. 基于 SIP 协议的社交服务平台的设计与实现 [D]. 成都: 电子科技大学, 2017:31-32.
 [4] 拜亚萌, 张燕玲. 大数据时代下高校毕业生就业网络“精准化”服务模式研究 [J]. 科技视界, 2019:6-7.