

基于文本挖掘的成都市电子信息产业中高端人才需求研究

胡庭宇, 李贵卿, 郭 彤

成都信息工程大学管理学院, 四川 成都

收稿日期: 2023年6月29日; 录用日期: 2023年7月7日; 发布日期: 2023年8月16日

摘 要

为更好的了解目前成都市电子信息产业中高端人才的现状, 本文从中高端人才的需求端出发。运用Python的数据抓取技术从智联招聘上搜索招聘数据, 对岗位类型、薪资区间、工作年限、学历、社会组织信息五方面绘制出该产业人才需求画像, 并应用关联规则中的Apriori算法分析其潜在联系。并得出目前技术研发类型岗位需求量最大; 中小企业和民营企业为招聘主力军, 同时大型企业和国企也有就业机会; 薪资区间与学历呈正比等结论。

关键词

成都市, 电子信息行业, 文本挖掘, 关联规则

Research on the Demand for Middle and High Technology Talents in Chengdu's Electronic Information Industry Based on Text Mining

Tingyu Hu, Guiqing Li, Tong Guo

School of Management, Chengdu University of Information Technology, Chengdu Sichuan

Received: Jun. 29th, 2023; accepted: Jul. 7th, 2023; published: Aug. 16th, 2023

Abstract

In order to better understand the current situation of middle and high technology talents in the

electronic information industry in Chengdu, this paper starts from the demand side of middle and high-end talents. Using Python's data crawling technology to search recruitment data from Zhao-pin.com, we draw a portrait of the demand for talents in this industry in five aspects: job type, salary range, working years, education, and social organization information, and apply the Apriori algorithm in association rules to analyze their potential connections. And we conclude that the current demand for technology R&D type positions is the largest; small and medium-sized enterprises and private enterprises are the main recruiting force, while large enterprises and state-owned enterprises also have employment opportunities; salary range is positively proportional to education, etc.

Keywords

Chengdu City, Electronic Information Industry, Text Mining, Association Rules

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

2022年美国工业与安全局对中国高端芯片制造管制升级,以限制中国包括半导体生产能力,超算力提升。美国的限制也让中国电子行业人才市场更为活跃。

成都早在2021年其电子行业的企业产值就达到了1300余亿元,其产业规模排在全国前十。成都电子行业发展在西部中开始较早,其发展区域汇集了郫都、高新、天府新区等地,在产业功能区中较优秀的企业包括了英特尔、华为、德州仪器等全球知名企业,同时也有中国电科、华大九天等国企入驻。在人才攻击上也有电子科技大学、四川大学、成都信息工程大学等院校的支持。同时成都政府为推行电子产业的发展,相继出台《关于促进电子信息产业高质量发展的实施意见》和《成都高新区关于支持集成电路涉及产业发展的若干政策》等政策作为支撑,意图打造集“芯-屏-端-软-智-网”为一体的电子信息产业体系。

本文从需求端对数据进行采集、处理、分析,以期更好的了解成都市电子行业中高端人才市场的需求现状。通过岗位类型、薪资、工作年限、学历、社会组织信息等当面绘出人才需求画像。

2. 研究现状

目前已有不少研究人员对这方面的内容进行了深入透彻的研究。郝素利等(2022)运用python技术对前程无忧、智联等招聘网站进行了数据采集,并结合TextRank算法、Apriori关联规则的方法对会计人才的整体需求以及岗位技能需求做出了分析[1]。赵玲(2019)运用集搜客爬取了第一季度成都地区人力资源行业的招聘信息,并对工作经验、学历要求、岗位需求等方面进行了分析说明[2]。支凤稳等(2023)利用网络爬虫、自然语言处理、社会网络分析等方法,对区域信息资源管理人才的招聘需求进行分析并尝试提出了其培养路径[3]。尹晶海等(2022)在搜集安卓人才的招聘信息后,利用Jieba包和TF-IDF算法对关键词进行了抽取,分析了其岗位名称,并用关联分析了不同岗位的收入[4]。谭昕章等(2021)利用Python三方采集软件对电商运营岗位进行采集,利用Excel和SPSS对数据进行了全方位的分析,以对该专业学生的职业规划培养提供一定意义的参考价值[5]。梅杰(2021)对网上搜集到的信息御用DTH-Apriori算对岗位信息、薪资报酬、企业规模等数据进行了关联分析,并通过TextRank算法以及TI-IDE获取了热门职位的关键词[6]。黄山等(2017)运用python语言抓取数据,并通过Jieba分词模块对岗位信息描述进行了中

文分词,在 SPSS 中使用 Apriori 算法,对学历、工作年限及其工作地进行了关联分析[7]。刘彩娜等(2023)利用 Java 版本的 Jieba 方法实现中文分词,通过中文分词、去停用词、结果降维、优化物流词库及分词得到智慧物流人才的需求关键词,并从业务知识、信息技术以及基本素质需求三个方面进行了分析[8]。

3. 数据采集与预处理

3.1. 数据采集

本文从招聘网站的用户数量及知名度考虑,最终选取从智联招聘网站中获取岗位数据。本文基于 Python 的爬虫技术爬取了成都相关数据,展示成都市目前电子行业中高端人才的需求情况。因中高端人才大多为研究生学历,故在数据获取当中,将筛选条件中行业条件设置为电子/半导体/集成电路,并依次对学历设置为本科、硕士、博士进行搜索。共搜集到数据 1360 条。

3.2. 数据预处理

数据预处理部分主要对部分错误和重复的信息进行了数据清洗。首先是采集中错误的信息。在信息采集过程中,部分信息产生错误,导致空白栏的产生,同时还有部分信息中的岗位名称与所需行业岗位不匹配,如:如体育产品经理、绿色低碳技术经理、总经理秘书等。对于这部分的信息予以直接删除。其次是重复的信息处理。采集中有招聘公司投放了相同的岗位,导致重复信息的出现。因此需要检查同一公司的发布岗位内容,去除掉重复多余的信息。最后对于部分缺失的信息,比如企业性质为空白,则在网上查询对应的信息予以补齐。

3.3. 数据整理

将所有信息做规范化处理,最终形成岗位名称、学历要求、薪资区间、薪资发放月数、工作年限、公司性质、公司规模 7 个关键字。尤其是薪资区间的标准化,介于原始数据的薪资区间分布较广且不规范,会出现 7 千~1.4 万、1.3 万~2.6 万等区间,故为规范化处理,统一对薪资数据将其上界和下界数据分开,取其均值,最终按照均值数对其进行区间的划分,同时去掉薪资为面议的数据。同时对工作年限、公司性质、公司规模等进行标准化操作,得到数据表 1。

Table 1. Recruitment information sheet

表 1. 招聘信息表

变量名称	中文变量名称	原数据展示
job_name	职位名称	嵌入式开发工程师
salary	薪资区间	1 万~2 万
salary_time	薪资发放月数	13
education	学历	本科
exp	工作经验	1~3 年
comp_qua	公司性质	民营
comp_scl	公司规模	100~299 人

4. 岗位数据分析

4.1. 岗位名称分析

通过 NVIVO 对岗位名称进行了词频分析,并分别对本科和研究生单独进行了词频分析。

通过表 2 和图 1 中对成都虽有岗位名称的分析来看, 工程师的占比始终是最高的, 占到了全部词汇的 18.39%, 其次是设计和销售, 分别为 2.50% 和 2.07%。目前从总的市场上来看, 对于设计和销售类人员的需求量更大。

Table 2. Job title word frequency analysis top 10 frequency word list

表 2. 岗位名称词频分析频数前 10 的单词表

总岗位名称分析前词频数 前 10 单词			本科类岗位名称分析前词频数 前 10 单词			研究生类岗位名称分析前词频数 前 10 单词		
单词	计数	加权百分比(%)	单词	计数	加权百分比(%)	单词	计数	加权百分比(%)
工程师	861	18.39	工程师	579	17.91	工程师	282	19.16
设计	117	2.50	销售	96	2.97	设计	81	5.50
销售	97	2.07	专员	86	2.66	ic	46	3.12
专员	89	1.90	助理	74	2.29	研发	40	2.72
软件	83	1.77	经理	64	1.98	算法	35	2.38
助理	82	1.75	软件	60	1.86	数字	34	2.31
研发	82	1.75	技术	59	1.82	高级	29	1.97
技术	75	1.60	硬件	57	1.76	芯片	26	1.77
经理	75	1.60	管理	46	1.42	开发	23	1.56
开发	68	1.45	开发	45	1.39	模拟	23	1.56

由此, 也可以对成都的岗位做一个区分。从岗位的级别上来看, 包含了业务人员和管理岗位, 其岗位有实习生、专员、助理、业务员、专家、技术员、经理、总经理、管理员、总监、董事长等职位。从岗位的类型来看, 又包含了技术类的职位和非技术的职位, 在词频统计中可以看出技术类的岗位需求量更大, 其占比更多。技术类的岗位包括了芯片、电子、电路、设计、开发、测试、工艺、算法、封装、检测、前端等岗位; 除却技术类岗位, 同时对非技术类的岗位也有一定的少量的需求, 出现行政、营销、销售、人事、秘书等词汇。按照技术的类型来分, 则有 IC、ASIC、CPU、EMU、FPGA、J10038、J10001 等。涉及的行业领域有汽车、自动化、通信、无人机等。

本文参考 2022 年成都市人才白皮书对集成电路行业的人才分类以及猎聘大数据报告中对集成电路/芯片人才的分类。其中 2022 年成都市人才白皮书中对集成电路分为了 ICC 设计, 封装、测试以及应用类。而在猎聘中分为了设计类和技术研发类, 跟设计相关的称为设计类岗位, 如数字前端工程师、模拟芯片设计工程师、IC 验证工程师等归为设计类工程师, 而与芯片制造、封装、测试、设备等芯片的相关的为技术研发类, 如半导体技术工程师、半导体工艺工程师、封装研发工程师列为生产类工程师。参考以上分类, 将职位分成四种类型, 即设计类岗位、技术研发类岗位、市场销售类岗位、职能型岗位。其中关键字包含设计以及 IC 验证等为设计类岗位, 涉及到芯片、制造、封装、测试、设备以及具体某种技术的为技术研发类岗位, 关键字中包含了市场、销售、项目经理等为市场销售类岗位, 而关键词为秘书、人事、行政、董事长、广告涉及类的为职能型岗位。

由此进行分类后发现, 招聘岗位类型最多的为技术研发类岗位, 其次是职能型、设计类以及市场销售类的岗位。可见电子信息行业的人才主体是围绕这技术研发类的人才展开的。

结合表 2、图 2 和图 3 来看, 可以看出研究生和本科岗位的变化。对于研究生学历的岗位来讲, 在招聘岗位名称中出现较多的词汇多为设计、算法、IC 芯片、模拟、开发等词汇, 同时词云图的展示中也



Figure 1. Word cloud of Chengdu job titles
图 1. 成都岗位名称词云图

可以看到如 FPGA、SOC 等的相关技术方向。结合实际招聘情况来看，有芯片工程师、算法工程师、模拟芯片设计工程师、数字 IC 工程师等岗位的招聘。而对于本科类的岗位来讲，则更多与软件、硬件、开发等联系在一起的，其次是是与销售、助理、专员、经理等代表岗位级别的词汇联系在一起。市场开拓与销售，始终是企业最核心的部分，事关企业未来的发展。本科生中的岗位多与销售工程师、项目经理、软件工程师、硬件工程师、研发助理工程师等挂钩。

综合两者比较来看，研究生的岗位招聘多为技术或者设计类的岗位。技术类的岗位多与芯片、算法、模拟、IC 相关，并且研究生岗位在技术方向上更为具体。本科生的岗位主要与技术研发类挂钩，除此之外销售和职能型的也偏向于招收本科生。本科生在技术方向上，多为硬件工程师和软件工程师。在猎聘大数据中对半导体/芯片人才的统计中，认为从招聘需求来看，芯片企业在研发生产和销售等领域的人才需求量是最大的。



Figure 2. Word cloud of postgraduate job titles
图 2. 研究生岗位名称词云图

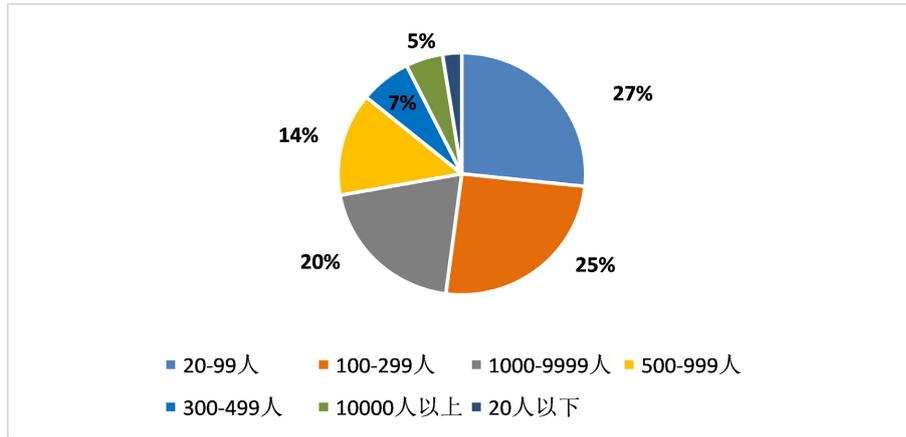


Figure 5. Statistical chart of company size as a percentage
图 5. 公司规模占比统计图

对本科学历的需求量更大，占到了 73%。研究生学历中，硕士学历占比较大为 26%。

从图 6 公司工作年限需求占比可知，总的来看 1~3 年工作年限要求更多，占到了 45.39%，其次是无限和 3~5 年分别为 24.93%和 14.34%。对比本科生和研究生的需求来看，对本科生工作年限的要求更为低，本科生工作经验要求 3 年以下的占到了 8 成，而对于研究生来讲，大部分招聘企业更需要其有经验的累计，虽然经验不限与一年以下与本科生的占比无异，但就是 1~3 年的占比为 24.44%，少于本科生 1~3 年的要求，而需要 3 年以上工作经验的公司比本科生的占比多了 20%左右。

在猎聘大数据对半导体芯片人才调查的报告中显示，目前芯片人才的招聘工作年限对比全行业来说要求更低，而学历对比全行业来讲，其对研究生学历的需求量是更大的。而就成都的形式来看，总体来说确实入门的门槛并不高，但是对于研究生的工作经验要求反而更高，也可以看出该行业入门简单，但是对于职业后期发展来讲，需要高智慧型人才，且培养周期长。

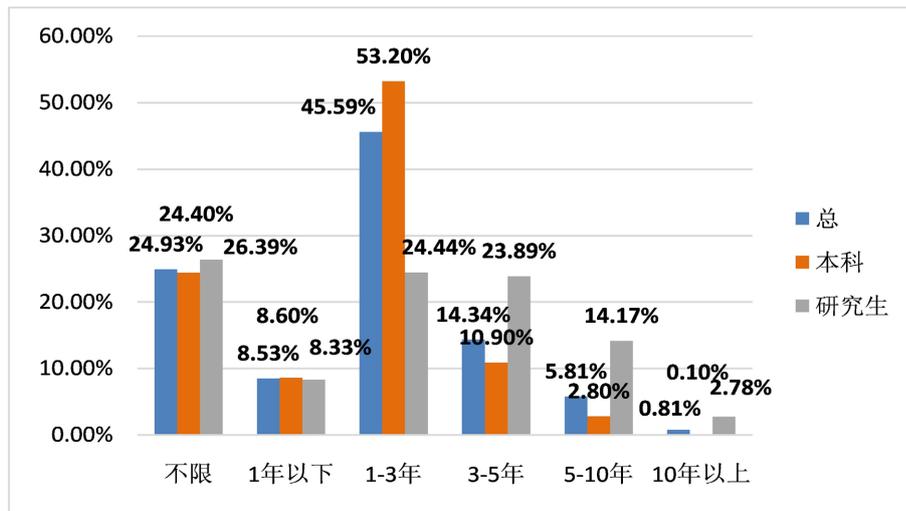


Figure 6. Statistical chart on the percentage of years of work required by companies
图 6. 公司工作年限需求占比统计图

4.4. 招聘薪资分析

根据薪资区间的统计图 7 可以看出目前成都市电子行业的薪资水平以及本科和研究生的薪资水平的差异。

从图7统计的薪资区间来看,目前在成都电子行业薪资水平多集中在5千~1万以及1万~2万这两个水平区间上,分别占比达到了50.63%和31.44%,占到了总体水平的八成。对于本科生来说,大多也集中与这个水平段上。而研究生的情况而言相对来说薪资水平会更高一些,很少有企业的薪资区间设定在1万以下。研究生1万~2万的薪资招聘占比达到了42.25%,并且在2~3万,3~4万,以及4万以上的占比也分别达到了29.86%,13.80%和10.99%。

考虑到部分企业会在年底有额外的薪水福利发放,故统计了薪资发放月数,大部分企业还是正常的12薪,但是有30%左右的企业选择了13薪和14薪,而15薪的占比有5.42%。选择发放16薪、17薪、18薪甚至更高的公司占比非常少。对于本科生来讲,大部分企业选择发放13~15薪,而研究生来讲,选择发放14~17薪作为激励手段的企业会更多一些。

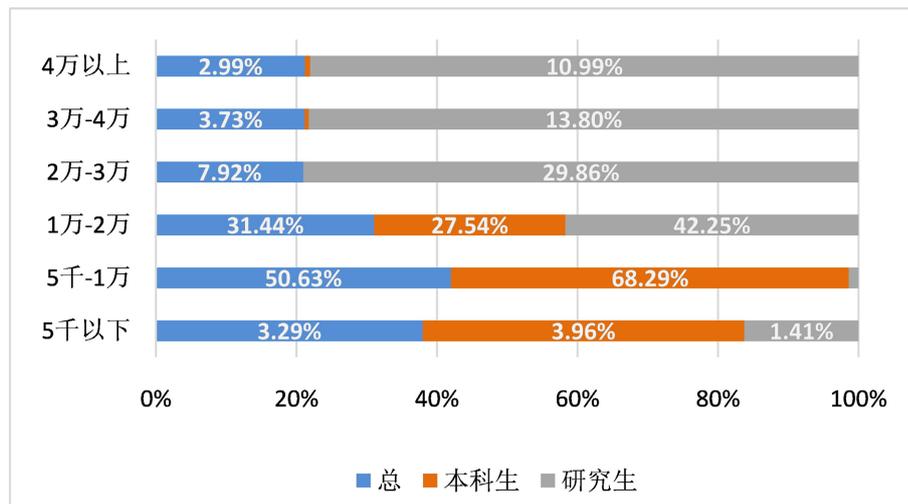


Figure 7. Comparison chart of salary range distribution
图7. 薪资区间分布对比图

4.5. 关联规则分析

为找出薪资区间、学历、工作年限、公司规模、岗位类型之间的潜在关联,本文应用关联规则中的Apriori算法,通过找出数据集中出现频率搞得子集模式,即找出频繁模式来对数据进行挖掘。通过操作得到表3的数据,其中最小支持度设置为20%,最小置信度设置为60%,提升筛选为大于1,当提升度大于1时,表明两者之间为强关联。

Table 3. Association rule sheet
表3. 关联规则表

序号	先导项	后继项	支持度	置信度	提升度
1	(5千~1万)	(本科)	50.26%	99.11%	1.35
2	(本科)	(5千~1万)	50.26%	68.47%	1.35
3	(1~3年)	(5千~1万)	27.64%	60.59%	1.19
4	(1万~2万)	(技术类)	20.75%	66.27%	1.21
5	(1~3年)	(本科)	39.25%	86.04%	1.17
6	(20~99人)	(本科)	23.52%	79.70%	1.09

Continued

7	(5千~1万, 1~3年)	(本科)	27.57%	99.73%	1.36
8	(本科, 1~3年)	(5千~1万)	27.57%	70.23%	1.38
9	(1~3年)	(5千~1万,本科)	27.57%	60.43%	1.20
10	(5千~1万, 技术类)	(本科)	25.24%	98.83%	1.35
11	(技术类, 本科)	(5千~1万)	25.24%	64.56%	1.27
12	(技术类, 1~3年)	(本科)	20.45%	81.98%	1.12

根据表 3, 可以解释这 12 条强关联规则, 如规则 1, 如果一个企业的工资为 5 千~1 万, 那么这家企业有 99% 的概率招聘人员为本科; 规则 2, 如果一个应聘者的学历为本科, 那么有 68% 的概率薪资区间在 5 千~1 万……规则 12, 当岗位类型为技术类岗位, 且要求工作经历在 1~3 年的岗位, 有 82% 的概率是本科生。

由此规则可以得到的结论为第一, 当应聘者学历为本科的时候, 其条件大概率为, 薪资水平在 5 千~1 万, 要求的工作年限是 1~3 年, 从事的职位多为技术类, 招聘单位的规模在 20~99 人的企业。第二, 招聘月薪在 1 万~2 万区间范围内的, 其从事的职位类型大概率是技术类的工作, 但是技术类薪资区间在 5 千~1 万的岗位, 多数是招收本科学历。除此之外, 如果降低最小支持度值到 10% 的时候, 会发现学历本科与职能型岗位也具有点概率的强关联; 而从事设计类岗位的人才, 大概率学历为硕士, 这一点与岗位和学历的词频分析是一致的。

5. 结论

从总体来看, 目前成都市电子信息产业企业对中高端人才招聘的岗位, 可分为技术研发类、设计类、市场销售类以及职能型。在这四种类型中目前对技术研发类人才的需求量最大。在技术类型中, 本科生多为硬件工程师、软件工程师、技术助理等, 研究生从事的职位大多有较为具体的技术方向, 岗位专精。设计类的岗位大多都要求学历是硕士或硕士以上。而技术研发类、市场销售类以及职能型的岗位大多对学历的要求为本科生即可。

在招聘的社会组织中, 民营企业、中小企业仍然是招聘的主力军, 但是国企、上市公司或者有一定规模的企业对人才的需求也不少。由此来看, 在成都想要进大公司其机会也是较多的。而对于规模较在 20~99 人的企业大部分在学历上的要求为本科。

而在对招聘者的学历、工作年限以及薪资区间来看。7 成对学历的要求为本科生, 总体来看对工作年限的要求集中在 1~3 年, 但是研究生对工作年限的需求会更高。从猎聘给到的大数据统计中, 电子信息行业对研究生的需求量是高于全行业的需求量, 而工作年限的需求量是低于全行业的水平。可见目前电子信息产业入门较为简单, 但是想要在产业里面有更好的发展, 深造和工作年限是个选择。在薪资区间中 5 千~1 万的薪资, 大多是对学历的需求是本科。而研究生的薪资区间在 1 万以上占比最大, 甚至有些企业的招聘薪资可以达到 4 万以上。从薪资发放月数来看, 企业对研究生也会提供更多的物质激励, 大部分企业会增加研究生的薪资发放月数, 增加薪资发放月数, 也是吸引人才就业的一种有效激励手段, 除此之外也可以引导学生就业以及专业的选择。由上述的薪资区分来看, 电子信息产业的薪资与学历也成正比的关系。设计类和技术类人才的薪资水平一般要高于职能型的人才。

基金项目

成都“高精尖缺”科技人才引进策略研究(2021-RK00-00106-ZF); 《数智化与思政化: 术道融合的

人力资源管理专业复合型人才培养体系创新实践》(G2021-1000);《新基建背景下 VR/AR/MR 技术赋能大学生创新创业能力培养研究》(项目编号: JYJG2022130)。

参考文献

- [1] 郝素利, 王瑞芳. 基于 Web 文本挖掘的会计人才需求分析[J]. 中国管理信息化, 2022, 25(19): 165-173.
- [2] 赵玲. 成都地区人力资源行业招聘需求对人才培养的启示——基于爬虫数据[J]. 财富时代, 2020(7): 109-110.
- [3] 支凤稳, 彭兆祺, 程镇朝, 马小琪. 京津冀地区信息资源管理学科人才需求分析[J/OL]. 情报理论与实践: 1-11. <http://kns.cnki.net/kcms/detail/11.1762.G3.20230320.0845.002.html>, 2023-05-18.
- [4] 尹晶海, 陈钰华. 基于文本挖掘的 Android 相关人才招聘岗位数据分析[J]. 电脑知识与技术, 2022, 18(21): 1-3.
- [5] 谭昕章, 杜晖. 电商运营岗位的社会需求及岗位素质分析——基于智联招聘网站数据[J]. 现代商贸工业, 2021, 42(3): 63-66.
- [6] 梅杰. 基于关联规则的网络招聘信息挖掘与分析[D]: [硕士学位论文]. 贵阳: 贵州大学, 2021.
- [7] 黄山, 刘学锋, 毛建华, 李银桥. 基于关联规则的大数据领域人才需求分析[J]. 工业控制计算机, 2017, 30(8): 78-80.
- [8] 刘彩娜, 路健, 封逍遥. 基于招聘信息的智慧物流人才需求文本挖掘[J]. 无线互联科技, 2023, 19(2): 138-140.