

# 招聘职位信息分析与可视化呈现

杨浩宇, 鄢田云, 张 帅, 刘春莉, 周 懿, 袁馨琪, 罗 艺, 杨 闰, 邓文丽, 郭 霞,  
柳敏烊, 李泽滔

成都信息工程大学应用数学学院, 四川 成都

收稿日期: 2024年11月4日; 录用日期: 2025年2月13日; 发布日期: 2025年2月25日

---

## 摘要

本文旨在借助大规模在线招聘职位信息的采集和分析, 深度挖掘招聘市场的动态趋势和需求特点, 为企业和求职者提供更全面的市场洞察和决策支持。首先, 从求职者角度和企业角度分别展开需求分析, 确定采用的技术方案及整体处理流程。然后, 完成数据采集和数据整理。从招聘网址以爬虫的方法获取原始数据集, 主要以Selenium与Edge WebDriver作为主要的网页爬取技术。数据爬取完成初始化网页浏览器、用户登录、信息提取、数据爬取、数据整合等工作。数据预处理完成数据去重、缺失值处理、数据拆分、数据整合、数据优化等任务。通过建立分类标准、应用自动化分类方法、手动审核与优化步骤完成数据分类流程。其次, 为了有效地捕捉到文本数据的深层结构和含义, 通过数据预处理、特征提取、模型建立、主题划分等流程, 结合TF-IDF技术进行分析, 对数据建立LDA主题模型, 通过特征提取与主题分析, 成功抽象出行业洞察。最后, 用Matplotlib库对数据进行可视化呈现, 完成基本数据可视化, 对技能、薪资趋势和地区差异等数据展开关联分析, 为招聘市场的动态变化提供了深入理解和全面分析。本文能为招聘市场的参与者提供深入洞察和一些实用指导。

---

## 关键词

爬虫技术, 数据挖掘, 机器学习, LDA主题模型, TF-IDF, 数据可视化

---

# Analysis and Visualization of Job Position Information

Haoyu Yang, Tianyun Yan, Shuai Zhang, Chunli Liu, Yi Zhou, Xinqi Yuan, Yi Luo, Run Yang,  
Wenli Deng, Xia Guo, Minyang Liu, Zetao Li

College of Applied Mathematics, Chengdu University of Information Technology, Chengdu Sichuan

Received: Nov. 4<sup>th</sup>, 2024; accepted: Feb. 13<sup>th</sup>, 2025; published: Feb. 25<sup>th</sup>, 2025

---

## Abstract

This thesis aims to explore dynamic trends and demand characteristics in the recruitment market

文章引用: 杨浩宇, 鄢田云, 张帅, 刘春莉, 周懿, 袁馨琪, 罗艺, 杨闰, 邓文丽, 郭霞, 柳敏烊, 李泽滔. 招聘职位信息分析与可视化呈现[J]. 软件工程与应用, 2025, 14(1): 8-16. DOI: 10.12677/sea.2025.141002

through extensive collection and analysis of online job postings. And it provides comprehensive market insights and decision support for businesses and job seekers. Initially, the demand analysis is conducted from the perspectives of both job seekers and businesses, the technical approach is determined, and the overall process flow is outlined. The subsequent study involves the completion of data collection and organization. Data acquisition from recruitment websites for job posting data employs web scraping techniques, primarily utilizing Selenium and Edge Web-driver. This process entails browser initialization, user authentication, information extraction, scraping, and subsequent data integration. Data preprocessing tasks include eliminating duplicates, handling missing values, splitting and integrating data, and data optimization. The data classification workflow is completed through the establishment of classification standards, the application of automated classification methods, subsequent manual review and optimization processes. Furthermore, In order to effectively capture the deep structure and meaning of textual data, an LDA topic model is established through data preprocessing, feature extraction, model establishment, and topic segmentation processes combined with TF-IDF analysis. By extracting features and analyzing topics, industry insights are successfully abstracted. Finally, data visualization is performed using the Matplotlib library. Basic data visualizations and correlation analysis of skills, salary trends, and regional differences are completed. Thus, this comprehensive analysis provides profound insights and practical guidance for participants in the recruitment market.

## Keywords

Web Scraping Technology, Data Mining, Machine Learning, LDA Topic Model, TF-IDF, Data Visualization

Copyright © 2025 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

在全球化的今天，互联网和信息技术发展如潮水般汹涌，深刻影响着社会的各个层面，尤其是在招聘这一关键环节。线下与线上各种招聘渠道和方式不断演变，使求职者面临更加激烈的竞争，这一变革也为求职者带来了前所未有的机遇与挑战。在海量招聘信息中，求职者需要筛选出最适合自己的职位，这一过程的复杂性和难度显著增加。因此，对招聘职位信息的精确分析与有效可视化推荐显得尤为重要 [1]。这不仅有助于求职者准确把握人才市场的动态趋势，优化求职策略，也能显著提升求职过程的效率与成果质量。

与此同时，随着技术的不断进步，大数据分析和人工智能的应用使得对海量招聘数据的处理和分析成为可能。企业通过深入分析这些数据，可以更清晰地理解行业内人才需求的变化，掌握招聘动向，从而制定出更加科学合理的人才招聘策略。此外，将分析结果以直观形式通过图表、图形等可视化手段展现，不仅加快了决策者对信息的理解与吸收，还提升了招聘流程的透明度，增强了求职者与企业之间的信任。

目前许多求职者和企业在招聘过程中依然面临信息不对称的问题，导致求职者难以准确把握市场需求，企业也难以有效识别合适的人才。这种信息的缺失和不对称在一定程度上制约了招聘效率和质量。

本研究旨在通过深入分析和直观可视化招聘职位信息，向求职者展示企业需求和优先考虑的技能，帮助他们更准确地了解市场动向和职业规划[2]。同时，研究结果也为企业招聘人员提供市场需求和竞争环境中的关键模式与趋势的洞察，使企业能够优化招聘策略，精准定位人才需求。这种双向透明的信息

流动不仅助力求职者调整求职意向和职业路径以适应未来的市场需求，也使企业在激烈的市场竞争中高效招揽合适人才，从而节省招聘成本，提高整体招聘效率和质量。

## 2. 需求分析

### 2.1. 求职者需求分析

求职者对招聘职位信息的需求日益精准化，具体体现在对职业信息准确性的要求上。求职者希望获取详尽且准确的职业信息，包括职位要求、薪资待遇和工作环境等方面。通过对招聘职位的分析与可视化，求职者能够更清晰地了解职业发展前景及行业趋势，从而做出更为明智的职业选择。

求职者对公司文化与薪酬待遇的关注也在不断增加。在选择工作时，他们更加注重公司文化和薪酬福利。通过数据分析与可视化呈现，求职者可以迅速了解公司的核心价值观、团队氛围以及相较于市场的薪资水平，从而更好地与自身的期望和能力相匹配。

### 2.2. 公司需求分析

在当今竞争激烈的人才市场中，企业对招聘职位分析与可视化的需求愈加迫切。

企业必须重视人才市场洞察的重要性，以便准确了解人才市场的供需状况及其发展趋势，从而制定有效的招聘策略。招聘职位分析能够帮助企业精准定位人才需求，预测行业发展趋势，并了解竞争对手的人才布局。

企业对招聘策略优化的需求也日益增强。基于数据分析的招聘策略能够显著提升招聘的效率与质量。通过对职位数据、人才流动情况及市场薪酬水平的深入分析，企业可以优化招聘流程，降低成本，并增强对人才的吸引力。

## 3. 招聘职位信息分析与可视化系统设计与实现

### 3.1. 技术架构

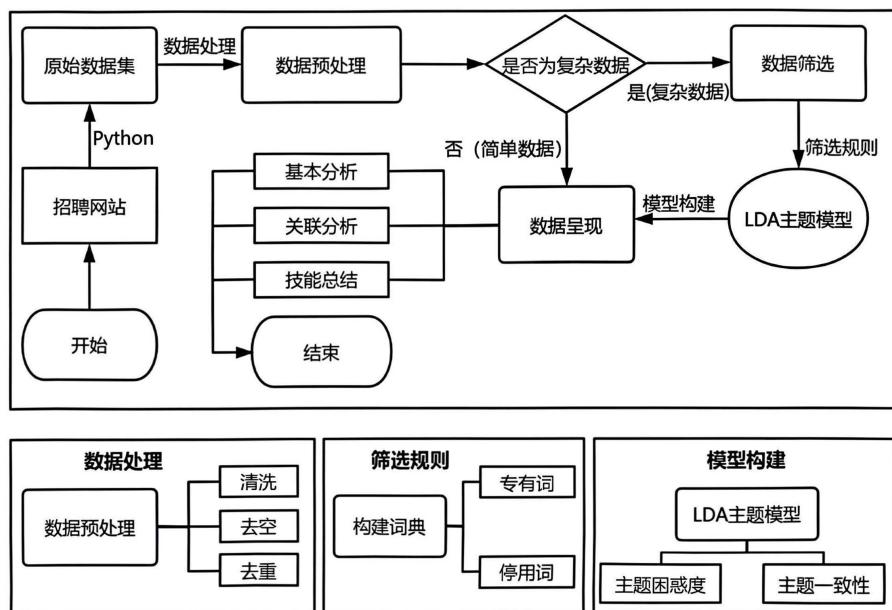
从招聘网址以爬虫的方法获取原始数据集，然后经过 Pandas 对数据集进行清洗、去空、去重。对处理后的数据进行判断，确认其数据内容是否复杂，对简单数据可以直接进行可视化呈现，对复杂数据以特定的筛选规则进行分词处理，然后对处理后的数据建立 LDA (Latent Dirichlet Allocation)主题模型，最后用 Matplotlib 等库对数据进行可视化呈现。具体研究流程如图 1 所示。

#### 1) 爬虫相关技术介绍

网页爬虫，也被称为网络蜘蛛或自动索引器，是一种自动化的程序，旨在从互联网上下载网页，从这些页面中提取信息。这些爬虫程序模拟用户浏览行为，通过系统地浏览网页来索引网上的信息，这对于数据挖掘、信息检索以及网络内容监测等领域至关重要。爬虫技术主要分为以下几种方法：第一，基于规则的爬虫依赖于解析 HTML (Hypertext Markup Language)源码，利用 BeautifulSoup 等工具来提取数据，这种方法适用于静态网页内容的提取。其次，使用网站提供的 API (Application Programming Interface) 进行数据爬取，这取决于网站是否提供了可供调用的 API，此方法对于获取结构化数据非常有效。再次，基于浏览器的爬虫，通过模拟浏览器行为(如 Selenium 和 Puppeteer 等)来解决 JavaScript 渲染页面的问题。每种方法都有其适用场景以及优缺点。

在本研究中，主要以 Selenium 与 Edge WebDriver 作为主要的网页爬取技术，以获取所需的网页信息 [3]。Selenium 是一个用于 Web 应用程序测试的工具，可以直接在浏览器中模拟用户操作，例如点击、填写表单等，同时可以获取页面的 HTML 内容。而 Edge WebDriver 是 Microsoft Edge 浏览器的一个 WebDriver 实现，它可以与 Selenium 结合使用，通过控制 Edge 浏览器来实现自动化的网页操作和信息获取。

通过 Selenium 与 Edge WebDriver 的结合, 程序能够实现对目标网站的自动化访问, 并且可以模拟用户的真实操作行为, 避免了一些网站针对爬虫的反爬措施。同时, 由于 Edge WebDriver 是 Edge 浏览器的官方支持工具, 因此在浏览器兼容性方面具有较好的稳定性和兼容性, 能够有效地应对网页结构的变化和动态加载内容的情况, 从而确保爬取的准确性和完整性。



**Figure 1.** Research flowchart  
图 1. 研究流程图

## 2) LDA 主题模型

LDA 主题模型是一种高效的概率统计模型, 它主要用于从大量文本数据中分析和发现隐含的主题结构。该模型基于一个假设, 即每个文档是由多个主题的混合构成, 而每一个主题又由一系列特定的词语表征。在实际应用中, LDA 模型通过推断文档的主题分布和各主题的词分布, 揭示文档集合中深藏的、不易直接察觉的主题信息, 从而为我们提供对文本集合深层次理解的途径。

使用 LDA 主题模型分析文本数据的过程中, 模型的应用带来了多方面的优势。首先, LDA 模型能够有效地识别出文本数据中的潜在结构, 这些结构作为文本的抽象表示, 显著提高了我们对文本内容核心特征的认识。此外, 这种模型还能实现数据维度的显著降低。通过将高维的文本数据转换为相对低维的主题分布, LDA 不仅简化了数据结构, 还为后续的数据处理和分析任务(如分类、聚类等)奠定了基础, 提升了这些任务的执行效率与准确性。

此外, 将文本数据通过 LDA 模型转换成固定数量的主题分布后, 可以更为便捷地将各种机器学习算法应用于这些数据上, 进一步增强了数据分析的效果并提高了其准确性和可解释性。值得一提的是, LDA 模型的适应性强, 不受特定文本长度或结构的限制, 使其能够灵活应对从短文本到长文章的各种规模和类型的文档集合, 这使其在各种应用场景下都表现出了优异的性能。

在本研究中, 将 LDA 主题模型应用于通过爬虫技术获取的招聘网站数据, 有助于我们从庞大的招聘信息中抽取出关键的职业领域和技能需求主题。这种深入的分析不仅揭示了就业市场的当前趋势, 而且为求职者与雇主之间的信息匹配提供了宝贵的洞见, 极大地促进了双方需求的精确对接。因此, 通过使用 LDA 主题模型, 就能够有效地提取职责和公司描述中的主题信息, 进而实现对文本数据的分类和可视

化分析。

### 3.2. 开发工具与主要库

本次研究所使用到的开发工具与主要库如表 1 所示。Pycharm 是用于编写 Python 代码的平台，本文采用 Python 版本为 3.11。Selenium 是一款用于 Web 自动化测试的工具，能够模拟用户操作。Edge WebDriver 是 Microsoft Edge 浏览器的自动化测试接口，用于与 Edge 浏览器进行交互。Matplotlib 是 Python 中常用的绘图库，用于数据可视化和图形展示。Jieba 是一款用于中文文本分词的库，能够将中文文本切分成词语。Scikit-learn (Sklearn) 是一款机器学习库，提供了各种机器学习算法和数据处理功能，用于构建和训练机器学习模型。Pandas 是一款用于数据处理和分析的工具，支持数据清洗、转换和分析操作。pyLDAvis 是一个用于交互式主题模型可视化的 Python 库，可帮助用户更好地理解文本数据中的主题结构。“gensim”是一个用于文本数据处理和主题建模的 Python 库，可以用于文本相似度计算、文档相似度计算等任务。在自然语言处理和文本分析中，“gensim”能够提供高效的工具和算法，支持主题模型的构建和应用。注意本小节所说的库仅指使用频率最多或最关键的库，并不包含所有的库。

**Table 1.** Development tools table

**表 1. 开发工具表**

序号	名称	用途
1	PyCharm	Python 开发平台
2	Selenium	Web 自动化测试工具，模拟用户操作
3	Edge WebDriver	Microsoft Edge 的自动化测试接口
4	Matplotlib	Python 绘图库，用于数据可视化
5	Jieba	中文分词库，用于将中文文本切分成词语
6	Sklearn	机器学习库，支持各种算法和数据处理，用于模型构建和预测
7	Pandas	数据处理工具，用于数据分析、清洗和转换
8	pyLDAvis	用于 LDA 主题模型可视化。
9	gensim	文本处理和主题建模

### 3.3. 数据采集与处理

#### 1) 爬虫对象

分析当前计算机行业的就业市场，针对不同技能需求和工作机会的变化，本次研究将集中在获取和分析拉勾网上发布的相关招聘信息。

本研究主要爬取以下信息：公司名称、职业名称、工作地点、薪资水平、能力要求、工作职责与公司描述等。

#### 2) 爬取流程

本研究的数据爬取流程涉及六个核心步骤：初始化网页浏览器、用户登录、信息提取、数据爬取、数据整合，以确保高效、系统地收集所需数据。

##### ① 初始化网页浏览器

使用 Selenium 库初始化一个 Edge WebDriver [4]，通过修改浏览器设置来避免被网站识别为自动化爬取行为。这一步骤关键在于配置浏览器以模拟真实用户的访问行为，从而减少被目标网站检测到的风

险。

### ② 用户登录

由于网站的招聘信息需要登录后才能访问, 研究中通过模拟填写用户名和密码的方式完成自动登录, 以此获取登录后才能浏览的职位信息。

### ③ 信息提取

在分析页面结构时, 可以注意到招聘信息通常包含在类名为“item\_10RTO”的标签内。于是, 为了提取招聘信息, 可以采用了一种基于CSS (Cascading Style Sheets)选择器的方法来选择包含所需信息的类名, 然后使用Beautiful Soup提供的方法来提取数据。

### ④ 数据爬取

研究选定聚焦于计算机相关的工作分类, 通过循环访问每个相关页面, 爬取每个页面上的职位列表。每一页的URL通过改变页码参数来实现对不同页面的访问。采用随机等待时间来模拟真实人类浏览行为, 以避免因访问频率过高而被网站封禁。

### ⑤ 数据存储

所有爬取到的数据被收集并存储在一个列表中, 每个职位的信息组成列表的一个元素。最终, 这个列表被转换成Pandas Data Frame, 并导出为CSV文件[5], 方便后续数据处理和分析。

### ⑥ 数据整合

由于不同计算机行业职位的需求和工作特点不同, 从而导致各个子领域之间存在显著差异, 单一职业分类下的数据可能不足以全面反映整个行业的就业状况。因此从不同的分类中分别进行相同步骤的爬取, 最后将多个文档进行汇总, 最终爬取获得原始数据集。部分原始数据集如图2所示。

A	B	C	D	E	F	G	H
1	公司	职业	地点	薪资	要求	职责	公司描述
2	信也科技	委外监理	上海·张江	15k-24k	经验3-5年 / 本科	科技金融 金融业	“上市公司; 上下班不打卡”
3	信也科技	机器人产品经理	上海·张江	25k-35k	经验3-5年 / 本科	资深产品经理 语音 科技金融	“上市公司; 上下班不打卡”
4	信也科技	黑盒测试高级工	上海·张江	18k-23k	经验1-3年 / 本科	自动化测试经验 iOS/Android	“上市公司; 上下班不打卡”
5	平安科技	387114-项目经理	深圳·南山区	15k-25k	经验1-3年 / 本科	云计算 项目管理	“五险一金, 节日福利, 绩效奖金, 定期体检”
6	法本信息	SAP BW模块顾问	西安·雁塔区	12k-16k	经验3-5年 / 大专	ERP HANA数据库 ABAP BW	“不出差, 长期稳定, 员工班车, 健身房, 大牛多, 前景好, 晋升快”
7	深信服科技	AI算法工程师	深圳·西丽	20k-40k	经验1-3年 / 本科	人工智能服务 信息安全 自动化	“绩效奖金、节日福利、带薪年假”
8	手付通	综合行政	深圳·福田区	8k-16k	经验1-3年 / 本科	软件服务   咨询	“公司氛围好”
9	载道科技	ui/ux设计师	广州·黄埔区	6k-8k	经验不限 / 本科	软件服务   咨询 移动端/API	“五险一金, 年终奖, 带薪年假, 节日福利”
10	南湖院	测试经理	北京·西苑	20k-40k	经验5-10年 / 硕士	自动化测试经验 测试工作	“团队氛围好”
11	百融云创	高级java工程师	北京·望京	25k-35k	经验5-10年 / 本科	科技金融 后端工程师	“团队氛围好”
12	清研兰亭	商务BD	北京·五道口	15k-29k	经验3-5年 / 本科	人工智能服务 市场 数据服务	“团队氛围融洽、扁平化管理、技术驱动”
13	远哲视讯	销售	上海·虹口区	8k-12k	经验1-3年 / 大专	软件服务   咨询 信息安全	“销售提成”
14	知衣科技	图像算法工程师	杭州·萧山区	20k-40k	经验3-5年 / 本科	人工智能服务 PyTorch 数据	“大牛团队, 周末双休”
15	51WORLD	技术美术(J1025)	上海·徐汇区	20k-40k	经验3-5年 / 本科	软件服务   咨询 人工智能	“行业大牛”
16	彩讯股份	产品经理	深圳·南山区	14k-20k	经验3-5年 / 本科	软件服务   咨询 TO B IT	“上市公司”

Figure 2. Partial raw data chart

图2. 部分原始数据图

### 3) 数据分类

在完成数据预处理后, 接下来还需要对职业名称数据列进行精确分类, 这对于后续的数据分析至关重要[6]。通过根据预设的标准分类, 能够将数据集中的数据条目归属到不同的职位类别中, 从而便于执行更详细和更具体的分析。

#### ① 建立分类标准

按照研究目的和数据集特性, 需要建立一套详尽的分类标准, 以确保后续的数据分类工作能够有效进行, 并保证分类的准确性和可靠性。例如: 需要根据职业名称中的关键词, 如“软件”、“数据”、“销售”等来定义“软件开发与工程”、“数据与分析”、“销售与市场营销”等职位的大类别。这一步

骤涉及对行业常识的理解，以及与研究目标的对齐。

### ② 应用自动化分类方法

利用 Python 编程语言中的 Pandas 库，编写了自动化的数据分类程序，该程序能够遍历数据集的每一条记录，并依据预先设定的关键词列表对职位进行分类。关键词列表以一种结构化的方式存储在字典中，程序会逐一检查职位名称，以确定其所属类别。这种自动化分类方法不仅极大地提高了数据处理的效率，还减轻了人工分类的负担，确保了分类过程的一致性和准确性。

### ③ 手动审核与优化

尽管自动化分类极大地提高了效率，但仍然要进行手动审核以确保分类的准确性。对于自动化程序难以准确处理的边缘案例，需要进行人工复查，并据此调整和优化关键词列表，以提高分类算法的准确性。

通过建立分类标准、应用自动化分类方法、手动审核与优化这三个步骤的有机结合，构成了一个完整的数据分类流程。这种综合性的方法不仅提高了数据处理的效率，还确保了数据分类的准确性和可靠性。分类完成后，每条数据记录都被明确归类到一个预定义的职位类别中。这种分类结果使得后续的分析能够更加有针对性。分类后的数据集，其中部分展示如图 3 所示。

	A	B	C	D	E	F	G	H	
1	公司	职业	类别	地点	最低薪资	最高薪资	经验要求	学历要求	职责
2	信也科技	委外监理	行政与管理	上海	15k	24k	经验3-5年	本科	科技金融
3	信也科技	机器人产品经理	硬件工程与电子技术	上海	25k	35k	经验3-5年	本科	资深产品经理
4	信也科技	黑盒测试高级工程师	硬件工程与电子技术	上海	18k	23k	经验1-3年	本科	自动化测试
5	平安科技	387114-项目经理	项目管理与运营	深圳	15k	25k	经验1-3年	本科	云计算 项目经理
6	法本信息	SAP BW模块顾问	法律与咨询	西安	12k	16k	经验3-5年	大专	ERP HANA
7	深信服科技	AI算法工程师	人工智能与算法	深圳	20k	40k	经验1-3年	本科	人工智能工程师
8	手付通	综合行政	行政与管理	深圳	8k	16k	经验1-3年	本科	软件服务
9	载道科技	ui/ux设计师	设计与创意	广州	6k	8k	经验不限	本科	软件服务
10	南湖院	测试经理	硬件工程与电子技术	北京	20k	40k	经验5-10年	硕士	自动化测试

Figure 3. Partial categorized dataset

图 3. 部分分类后的数据集

### 4) LDA 主题模型

在处理复杂和多样化的职责描述文本时，传统的文本分析方法如词频统计或简单分类往往无法有效地捕捉到文本数据的深层结构和含义。这些方法可能忽略了文档中词语间的隐含关系和主题连贯性，限制了分析的深度和广度。为了克服这些限制，可以选择使用 Latent Dirichlet Allocation (LDA) 主题模型。

LDA 主题模型是一种强大的概率图模型[7] [8]，它可以从大量文档中自动发现隐含的主题，进而将文本内容以主题的形式表示。这种表示不仅能揭示文档集合中的普遍主题，还能显示单个文档中主题的具体分布。通过使用 LDA 模型，可以更好地理解和抽象复杂的文本数据，从而提取出更为丰富和有用的信息。

### 5) 数据预处理

在进行文本分析时，需要通过去除停用词、构建专用词典以及分词处理来进行数据预处理。停用词通常包括那些在文本中频繁出现但并不携带具体信息的词汇，例如：“的”、“是”等，这些词汇的存在可能会干扰对重要主题的提取。

为了更准确地处理和分析文本数据，我们还需要构建一个字典列表来定义特定的复合词或专有名词，如“五险一金”。这样做可以避免在文本处理过程中将这些复合词错误地拆分为单个词，从而保留其完整意义并提高分析的准确性，避免处理文本数据时生成意义不明的词语。

分词处理则利用“jieba”库进行中文文本的分词处理[9][10]。“jieba”是一种高效的中文分词工具，通过使用自定义词典和添加专用词来提升分词准确性。“jieba”支持多种分词模式，如精确模式、全模式和搜索引擎模式，可以根据实际需求选择合适的模式。

#### 6) 特征提取

在完成文本的预处理之后，下一步工作是将处理过的文本数据转换为模型能够理解和处理的特征形式。对于 LDA 主题模型来说，需要将文本转换为数学上可处理的向量形式，这一步骤至关重要，因为它直接影响到模型的性能和最终的分析结果。

##### ① 常见的特征表示法

特征提取的常用方法包括词袋模型(Bag of Words, BoW)和 TF-IDF 模型[11]。词袋模型通过统计每个单词在文档中出现的次数来表示文档，而不考虑词序和语法规则，这种方法简单但有效。与之相比，TF-IDF 模型不仅考虑了词在文档中的频率，还考虑了其在整个语料库中的分布，帮助我们区分哪些词是重要的，哪些词是常见的但不具信息量的。

##### ② 应用 TF-IDF 向量化

在本研究中，选择使用“sklearn”库的 TF-IDF 向量化方式来转换文本数据。TF-IDF 通过减少常见词的重要性并增加罕见词的重要性，提供了一种更加精确的特征表示方式，为 LDA 模型提供了更为稳定和有区分力的输入数据。通过这种转换，每个文本被编码成一个具有固定长度的向量，其中的每个元素对应语料库中一个特定的词，并通过 TF-IDF 分数表示这个词对文档的贡献。

## 4. 结论

在本文中，通过应用 LDA 主题模型，研究对大量复杂的招聘信息进行了分析，成功抽取了市场上的主要职业领域和技能需求。例如大城市，北京、上海和深圳对高科技职位的需求远超其他地区，这一趋势突显了先进技术职位集中在经济发达的城市区域。此外，研究发现，随着求职者的学历和经验的增长，薪资水平通常呈现出相应的上升趋势，特别是在科技和金融服务行业。

分析还表明，跨领域能力，如项目管理和多语言能力，以及对特定技术的熟练掌握，如机器学习和大数据分析，在招聘市场中具有高度需求。并且发现对这些能力的需求横跨多个行业，不限于技术行业。例如，在市场营销和客户关系管理职位中，数据分析技能已成为一个重要的加分项。

此外，随着远程工作模式的普及，地理位置对职业选择的影响正在减少，但对技术设备和网络连接的要求却相应增加。这一变化使得许多求职者有机会从事以往由于地理位置限制而无法从事的工作，进一步增加了职场的多样性和竞争性。

## 致 谢

成都信息工程大学应用数学学院鄢田云博士/副教授是本论文的指导教师和通讯教师，感谢鄢老师对本论文涉及的大创项目系统开发、论文撰写的悉心指导和修改建议，感谢成都信息工程大学教务处对本论文成果来源的大学生创新创业训练计划项目及论文版面费支持。

## 基金项目

四川省大学生创新创业训练计划项目(S202410621115)，成都信息工程大学创新创业训练计划项目(202410621323)，四川省大学生创新创业训练计划项目(S202410621117)，成都信息工程大学创新创业训练计划项目(202410621326)。成都信息工程大学 2024 年本科教学工程项目“数智项目”(在线实践平台数据可视化的数智实验教学模式创新)。

## 参考文献

- [1] Xin, L., Zhou, B. and Liu, P. (2024) Position Information Visualization Analysis and Personalized Recommendation Based on Ant Colony. *ICST Transactions on Scalable Information Systems*, **11**, 1-6. <https://doi.org/10.4108/eetsis.5061>
- [2] 刘一, 王跟成. 基于 Python 的就业趋势可视化分析系统[J]. 信息与电脑(理论版), 2021, 33(5): 99-101.
- [3] 刘静, 王凤, 孟星, 等. Python 在数据可视化中的应用案例分析[J]. 电子技术, 2023, 52(5): 391-393.
- [4] 蔡文乐, 周晴晴, 刘玉婷, 秦立静. 基于 Python 爬虫的豆瓣电影影评数据可视化分析[J]. 现代信息科技, 2021, 5(18): 86-89+93.
- [5] 汤飞弘. 基于 Python 爬虫的招聘信息数据可视化分析[J]. 软件, 2023, 44(1): 176-179.
- [6] 汤洋, 汤敏倩. 网络招聘信息中职业类型与专业领域的情报分析[J]. 情报杂志, 2017, 36(6): 72-77.
- [7] 施乾坤. 基于 LDA 模型的文本主题挖掘和文本静态可视化的研究[D]: [硕士学位论文]. 南宁: 广西大学, 2014.
- [8] Chen, Y.C. and Pan, R.J. (2022) Research on Data Analysis and Visualization of Recruitment Positions Based on Text Mining. *Advances in Multimedia*, **22**, 77-79.
- [9] 杨静. 基于文本挖掘的网络招聘信息分析[D]: [硕士学位论文]. 济南: 山东师范大学, 2019.
- [10] 田书丽. 互联网招聘数据分析与可视化系统设计与实现[D]: [硕士学位论文]. 重庆: 西南大学, 2022.
- [11] Aizawa, A. (2003) An Information-Theoretic Perspective of tf-idf Measures. *Information Processing & Management*, **39**, 45-65. [https://doi.org/10.1016/s0306-4573\(02\)00021-3](https://doi.org/10.1016/s0306-4573(02)00021-3)