



内蒙古师范大学计算机科学技术学院

毕业设计（论文）开题报告

题 目： 基于 Python 的求职招聘信息数据分析

专 业 计算机科学与技术（非师范）

学 生 项斐

学 号 20161102932

指导教师 张丽萍

日 期 2019/11/25

计算机科学技术学院制

说 明

1. 课题来源及研究的目的和意义

经过大学三年的学习，我即将步入社会，面临着找工作的问题。我发现不能简单快捷的了解社会上对人力的需求。所以我想设计一个可以简单明了的展示不同城市对不同职位，不同人员的需求，帮助求职者可以更好的制定未来的方向，帮助创业者了解城市情况以及接下来的发展方向，确定自己创业计划。不论是求职就业还是自主创业，对一个城市了解时必不可少的，对于一个人城市的消费生活水平，不同职位薪资的高低都是需要你了解的东西。而对于现在的我来说获取这些信息不仅比较困难而且十分的繁琐。所以我设计本产品通过对招聘信息分析来实现以上的问题。

2. 国内外在该方向的研究现状及分析

通过对网上的求职招聘软件的了解，暂时没有发现具有该功能的产品。现有产品仅是对不同企业的招聘信息进行发布，求职者可以通过一定的条件对自己希望的岗位进行大体的筛选。也可以对不同的职位在社区论坛里与其他求职者进行交流。信息共享是当下求职招聘软件的主要功能。我们需要对筛选出来的信息进行再次的筛选分析，才能了解该职业平均薪资以及招聘人数等信息。我们需要进行大量的分析筛选等工作。

3. 主要研究内容

本产品通过对网上发布的招聘信息进行统计分析，将以下信息进行展示。

- (1) 全国各省招聘人数；
- (2) 各省各城市招聘人数；
- (3) 各省对不同职业的招聘人数；
- (4) 各城市对不同职业的招聘人数；
- (5) 各省各城市对不同职位的招聘人数；
- (6) 各个职业在不同城市的平均薪资

可以对以下人群有所帮助：

- (1) 填报大学志愿时有喜欢的专业不知道报考哪里的考生，可以帮助他们

选择该专业招聘人数多，薪资高的城市。

(2) 填报大学志愿时有喜欢向往的城市不知道报考什么专业的考生，可以帮助他们选择该城市招聘人数多的专业。

(3) 各级的毕业生，帮助他们找到适合自己专业的工作。

(4) 想要回到自己家乡或者特定城市的创业者、求职者，帮助他们了解该城市的详细信息。

(5) 想要更换自己现在工作的人，可以给予他们未来发展的方向。

(6) 政府工作人员，可以对他们的统计工作体现一定的参考价值。

对于不同的人有不同的帮助，通过本产品也可以了解一个城市的发展水平与它未来几年的发展方向，对于将要定居或者已经定居的人有所帮助。通过对展示信息的了解我们可以更好的选择自己将来的职业。

4. 研究方案

程序用 python 语言进行编写，通过爬虫获取网上的招聘信息存储与数据库中，通过数据的分析将最终的结果以图的形式展示在程序的界面上。

产品的设计需要以下的技术支持：

- (1) python 语言
 - (2) 基于 python 的网络爬虫
 - (3) 基于 python 的数据分析
 - (4) 基于 Photoshop 的图片处理
 - (5) 数据库
- 爬虫：网络爬虫是一个自动提取网页的程序，它为搜索引擎从万维网上下载网页，是搜索引擎的重要组成。所有被爬虫抓取的网页将会被系统存贮，进行一定的分析、过滤，并建立索引，以便之后的查询和检索。
 - 数据分析：数据分析指用适当的统计、分析方法对收集来的大量数据进行分析，将它们加以汇总和理解并消化，以求最大化地开发数据的功能，发挥数据的作用。数据分析是为了提取有用信息和形成结论而对数据加以详细研究和概括总结的过程。
 - 数据库：数据库是“按照数据结构来组织、存储和管理数据的仓库”。是一个长期存储在计算机内的、有组织的、有共享的、统一管理的数据集合。数据库是以一定方式储存在一起、能与多个用户共享、具有尽可能小的冗余度、与应用程序彼此独立的数据集合，可视为电子化的文件柜——存储电子文件的处所，用户可以对文件中的数据进行新增、查询、更新、删除等操作。

5. 进度安排，预期达到的目标

- (1) 2019/9 —— 2019/12 学习相关知识，软件设计的工作
- (2) 2019/12 —— 2020/2 完成毕业设计
- (3) 2020/2 —— 2020/3 完成毕业论文初稿

通过上述的三个阶段将基本实现程序预期的所有功能。

6. 课题已具备和所需的条件、经费

对于招聘信息各个网站都是公开的信息，所以对招聘信息的爬取较为简单而且产品为纯软件产品不涉及硬件。

7. 研究过程中可能遇到的困难和问题，解决的措施

在软件的实现过程中可能存在以下问题

- (1) 爬取的数据不全面，招聘信息集中在个别省市无法进行全面的分析
- (2) 信息过于杂乱，对于职位的描述不详细。
- (3) 对于信息的分类不准确

8. 主要参考文献

(1) 熊畅. 基于 Python 爬虫技术的网页数据抓取与分析研究[J]. 数字技术与应用, 2017(09):45-46.

(2) 王芳. 基于 Python 的招聘网站信息爬取与数据分析[J]. 信息技术与网络安全, 2019(8).

(3) 赵丽娜, 李伟, 康犇, et al. 基于 Python 爬虫的借阅数据获取[J]. 北华航天工业学院学报, 2018(4):61-62.

(4) 刘熠. 基于 python 的 Web 数据挖掘技术研究 with 实现[D].

(5) 张喜红. 基于 Python 的中药材价格信息爬虫的设计[J]. 枣庄学院学报, 2019, 36(02):72-77.

(6) 聂晶. Python 在大数据挖掘和分析中的应用优势[J]. 广西民族大学学报: 自然科学版, 2018, 24(1): 76-79.

(7) 杨凯利, 山美娟. 基于 Python 的数据可视化[J]. 现代信息科技, 2019, 3(05):38-39+42.

(8) 严婷, 文欣秀, 赵嘉豪, et al. 基于 Python 的可视化数据分析平台设计与实现[J]. 计算机时代, 2017(12):58-60.

- (9)胡非, 韩永辉, 许超, et al. 基于Python的可视化工具研究与应用[J]. 冶金丛刊, 2018(13).
- (10) 牛小宝. 基于MySQL的云数据库设计与实现[D]. 2016.