

石嘴山市空气质量数据分析与可视化

吴 燕

(宁夏理工学院, 宁夏 753000)

摘 要 随着社会快速发展、人口数量不断增长, 空气质量问题越来越突出。本文以宁夏石嘴山地区空气质量为研究内容, 通过 Python 语言的爬虫技术完成该地区近八年来的历史数据的获取, 并对数据进行清洗、处理, 最后通过 HTML 语言、JS 语言、Echarts 等技术实现对数据的加载和可视化等功能。完成后的系统, 界面友好、数据分析准确度高、统计功能较完备、分析结果展示人性化, 为石嘴山地区的环境保护提供一系列决策依据。

关键词 空气质量 数据爬取 可视化

中图分类号: X51

文献标识码: A

文章编号: 1007-0745(2021)10-0026-03

1 研究背景及意义

空气质量与人们的生活和健康息息相关, 不论在室内还是室外, 人们的生活都离不开空气。我国自 90 年代以来, 已建成 5000 多个环境空气质量监测点, 其中涉及到的有从县、到市、到省最后到国家。从监测功能上讲, 涵盖城市环境空气质量监测、区域环境空气质量监测、背景环境空气质量监测, 还包括温室气体、酸雨、沙尘影响空气质量监测等^[1]。同时基于物联网技术、互联网、网络爬取技术等应用将不同类型、不同地域的监测点的数据实时汇总至国家气象局并进行实时分析和统计, 数据对民众开放。

宁夏石嘴山市是传统的资源依赖型城市, 形成了以煤炭、冶金、机械、化工、陶瓷等多类非常齐全的重工业体系。石嘴山市地处宁夏回族自治区最北端, 紧靠内蒙古地区。蒸发强烈、空气干燥, 年平均降水量不到蒸发量的十分之一。多年来快速的工业化发展进程, 虽然带来了经济体量的提升和该地区人们物质生活质量的提升, 但同时也带来了严重的空气污染问题。

基于当前石嘴山地区的空气质量监测已经较为完整的情况, 本文通过对该地区自 2013 年至今的所有空气质量数据、境内企业相关数据、城市周边环境数据等进行数据获取和数据分析。选取合适的数据分析方法和可视化技术, 设计具有实时数据、历史数据、PM2.5 浓度分析、AQI 指数分析的空气质量分析系统, 该系统能够在一定程度上具有自动化、网络化等特点。通过对影响空气质量的各种因子(如 PM2.5、SO₂、NO₂、CO、O₃)的数据分析结果为政府部门制订空气质量提升政策提供依据。

2 需求分析

2.1 对石嘴山地区空气质量数据进行实时采集, 并在此基础上进行相关数据分析

通过获取石嘴山地区空气质量实时数据, 分别采用图形 UI、监测点数据列表的方式展示 24 小时内的 AQI 指数趋势。

2.2 基于历史数据进行相关分析与结果展示

数据统计分析模块采用多样化的图表展示, 根据获取数据的内容, 统计分析模块包括对历年数据的整体统计(可

以使用 PM2.5 或 AQI 指数), 对比多年来的数据均值, 以直观对比年度数据。

2.3 绘制 2021 年以来 PM2.5 数据折线图, 直观展示空气质量走势

为更加直观的了解到历年来空气质量高点变化情况, 对历年数据中 AQI 指数 >50 的数值进行统计展示。

3 系统可行性分析

3.1 经济可行性

系统设计开发对硬件要求需求不高, 所需开发工具均为开源, 无需过多支出, 因此在开发阶段实现了经济可行。

3.2 技术可行性

系统实现过程所需相关技术主要涉及到的有 Python 语言、CSS 和成熟的网页解析技术以及数据库等技术^[2], 均为市场成熟技术且被开发者熟练掌握, 因此实现了技术可行。

3.3 操作可行性

该系统布局简单、操作简便、互动性强, 便于用户使用, 因此实现了操作可行。

4 系统总体设计

石嘴山市空气质量分析系统包括实时数据可视化管理模块、历史数据可视化管理模块两大模块, 每个模块下又包含若干子模块, 具体内容如下图 1 所示。

4.1 实时数据分析与展示模块设计

该模块分为数据获取、图形化数据展示、AQI 数据可视化三个子模块。

4.2 历史数据分析与展示模块设计

该模块分为历史数据列表展示、历史数据分析、历史数据可视化三个子模块。

5 系统数据库设计

根据系统需求分析, 本文所实现的系统在满足功能设计的基础上, 数据库包含实时数据管理表、历史数据管理表。

5.1 数据库 E-R 图设计

空气质量主要管理实体有城市、空气质量、监测站点等实体, 城市信息实体的属性有 ID、名称。空气质量实体的属性有

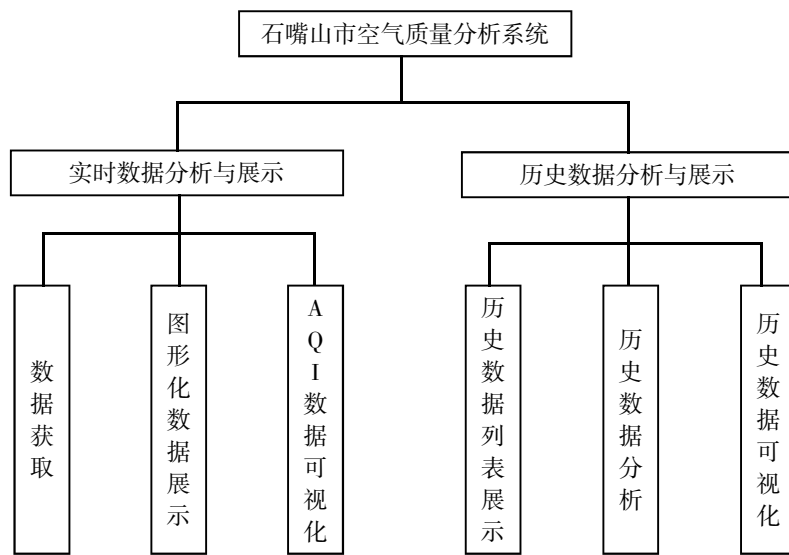


图 1 系统总体结构图

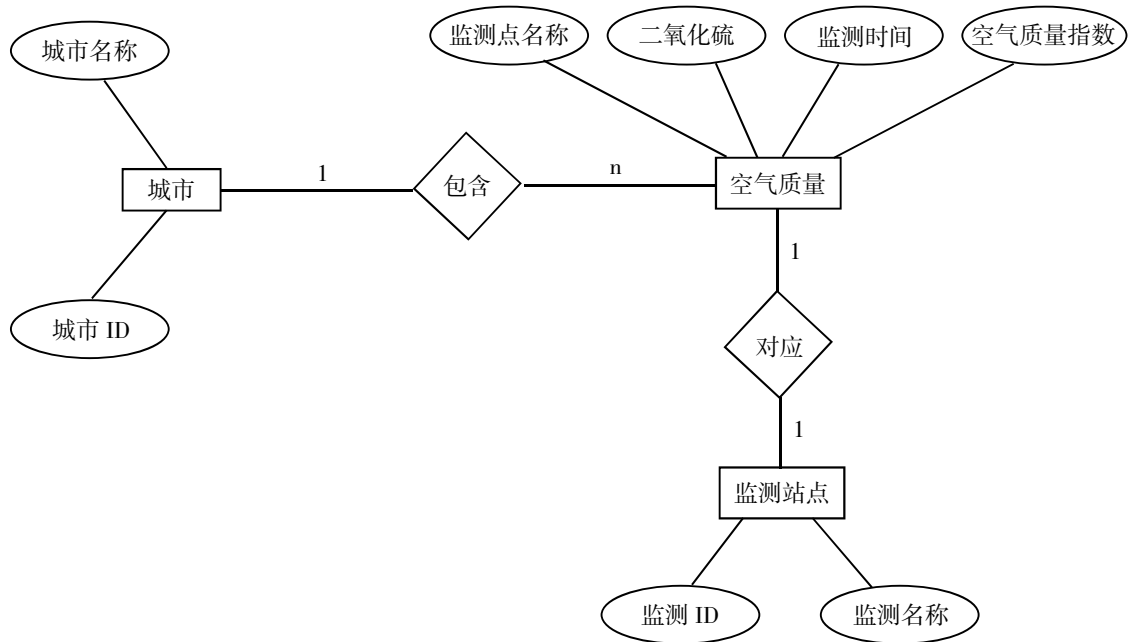


图 2 系统 E-R 图

二氧化硫、监测站点名称、监测时间、空气质量指数、AQI 等信息。监测站实体包括监测站名称、监测 ID 等信息。相关 E-R 图，如图 2 所示。

5.2 数据库逻辑结构设计

数据库逻辑结构设计就是把概念结构设计阶段完成的实体-关系图转化为与选用数据库管理系统产品所支持的数据模型相符合的逻辑结构。本系统数据库包含实时空气质量表与历史空气质量表。

6 基于网络爬虫与 API 的数据获取实现

通过网络爬取来获取网页中包含的数据，从而获取历史空气质量。通过开发者工具分析网页结构，所获取的数据是以半结构化 (HTML) 的格式存在于该网站，主要获取的字段包括日期、质量等级、AQI 指数、当天 AQI 排名、

PM2.5、二氧化硫、二氧化氮、一氧化碳、臭氧等。

通过 API (应用程序编程接口) 可以更为便捷地获取目标地区的空气质量数据，如 AQI 指数、空气质量等级等。该接口调用由 Python 的请求库访问，一般为 get 或 post 请求^[3]。请求返回结果以 JSON 格式解析，经提取数据后存入到数据库中。

7 数据分析过程及结果

7.1 历史空气数据的获取

空气质量数据的获取通过爬取目标网站“天气网”中的石嘴山市所有发布的空气质量监测的数据，存储于 MySQL 数据库中，通过 JSON 转换等加载到前台界面中。

7.2 历史空气质量数据列表化展示

该列表支持用户自动设置界面每次展示多少条数据、

界面上翻和下翻的后台服务端分页的功能。

7.3 历史空气质量数据可视化展示

数据统计分析模块采用多样化的图表展示,帮助用户直观了解过去以及现在空气质量指数以及PM2.5的变化情况及趋势^[4]。

根据本系统的数据获取的内容,数据统计分析模块应包括对历年来获取的数据整体统计(可使用PM2.5或AQI指数),并通过对比多年来的数据均值,以直观查看各年度数据。

7.4 数据分析结论

根据均值对比图统计,可知:2014年到2017年底,石嘴山市的PM2.5浓度一直较高。结合相关资料调查,这是由于石嘴山市在2018年前推动大气污染治理还未取得较好效果,境内以煤炭为主的生产加工企业较多。从2018年开始PM2.5浓度明显下降,说明政府以及民众的重视度提高,采取很大力度的措施得到显著的效果。但是2020年PM2.5又有了上升的趋势,是由于静稳天气的影响使得一月份PM2.5有升高的趋势,从而影响了整年度的PM2.5值。

从2021年变化走势图可知:2021年PM2.5浓度总体有所改善,仅有个别几天有明显上升的趋势。以3月15日为例,石嘴山市出现大范围沙尘暴天气,严重是影响空气质量。

结论:对比2013年到2021年3月份空气质量指数数据显示,以AQI指数为衡量标准,石嘴山市2018年以前的空气质量大于50的天气比较密集也比较多,从2018年以后

污染情况逐渐改善。

8 结语

随着人们对空气污染问题的重视度越来越高,空气质量在逐渐改善,基于该系统的实现,相信在以后的发展中,空气监测系统会越来越完善,会通过历史数据作出分析、预测,为监测人员提供依据。

总体而言,系统功能完备、界面交互性强,具有非常好的社会价值和使用意义。通过Bootstrap、CSS使界面更加美观,并减少开发所需时间,提高开发效率。随着人们对空气污染问题的重视度越来越高,空气质量在逐渐改善,基于该系统的实现,相信在以后的发展中,空气监测系统会越来越完善,会通过历史数据作出分析、预测,为监测人员提供依据。

参考文献:

- [1] 刘姝童,刘诗娇,夏鸣,梁钰薇,姚婷玮.天津机场运行对区域空气质量影响评估研究[J].科技风,2021(09):141-144.
- [2] 司桂静,崔向前,邹坤,戚昭岳.基于Python和Django实现城域网流量统计监控系统[J].山东通信技术,2020,40(01):26-27.
- [3] 白昌盛.基于Django的Python Web开发[J].信息与电脑(理论版),2019,31(24):37-40.
- [4] 李锦华,孙萌,陈缘,吕海波.2014-2020年渭南市雾霾发展状况分析[J].科学咨询(科技·管理),2021(03):70-71.

(上接第5页)

3.3 基于模型驱动的高铁设备自动测量

在完成系统标定后,通过现场测量可建立被测高铁部件与机器人坐标系之间方位关系,再结合高铁被测目标的三维模型或先验测量数据信息,可驱动机器人携带经纬仪到达目标位置,实现对被测目标的粗准直,再通过视觉引导准直技术实现被测目标的精确准直,如图6所示。通过对每个被测目标准直结果的坐标系统一,可实现不同被测目标坐标系间的相对姿态确定,从而实现最终测量。

该技术包含两个核心环节,一个是利用已知的被测目标在高铁坐标系下的先验位置信息,结合系统的标定结果,推算出经纬仪要移动到的目标位置及对应的水平俯仰角,从而实现对被测目标的准直;第二环节是经纬仪完成对被测目标的精确准直后,需要结合各系统间的标定结果将经纬仪在每个位置自身坐标系下的结果统一到全局坐标系下,而且尽可能减少坐标系传递误差,以保证最终高精度测量结果。

4 系统应用

相关技术先后应用于我国CRH5、CRH2-380、CRH380C(L)、CRH380D(L)、CR400AF等高铁型号中,完成了大重量部件装配,为新一代时速600km/h重要项目建设提供了重要支持,并由于技术改进,节约大量高铁研制经费,在支持国家战略建设的同时,也带来了可观的经济效益。机

器人自动测量系统可实现更高姿态测量精度,测量效率得到提高,同时测量人员进一步减少。

5 结论与展望

本研究通过自主攻关突破了机器人柔性控制、机器人高精度自动化姿态测量等关键技术,形成了面向高铁装配测试的机器人系列产品并在各型号平台应用,有效解决了高铁生产装配环节中面临的大质量设备安装柔性差精度低、装配测试自动化水平及效率低等难题。

未来将面向新一代高铁研制量增加的需求,重点突破机器人人机共融协作、仿人双臂机器人、自动化测量测试机器人等关键技术,建立起高度自动化、智能化的高铁总装测试生产模式。

参考文献:

- [1] 张立建,胡瑞钦,易旺民.基于六维力传感器的工业机器人末端负载受力感知研究[J].自动化学报,2017,43(03):439-447.
- [2] 喻洋,王耀兵,魏世民,马如奇,唐玲.基于柔顺控制的机器人装配技术[J].北京邮电大学学报,2020,43(04):1-6.
- [3] 汤海洋,纪柱,李论.基于力反馈牵引力导引的机器人辅助装配技术研究[J].制造业自动化,2021,43(03):9-13.
- [4] 黄心汉.微装配机器人:关键技术、发展与应用[J].智能系统学报,2020,15(03):413-424.