

基于爬虫技术的空气质量信息 可视化系统设计与实现

吴蔚然 彭博 欧必杰

南宁学院人工智能学院, 南宁, 530200

摘要 从多个角度对空气质量可视化系统的各项需求进行拟合, 开发一套国内空气质量信息可视化系统。系统由实时AQI数据图表显示模块、历史AQI数据图表自定义模块、数据获取模块组成。系统研发采用前后端分离的技术模型, 前端框架使用Vue, 后端框架使用Spring Boot。系统的空气质量数据获取是利用Python爬虫技术来获取, 通过ECharts实现对空气质量污染数进行图表可视化的展示, 并且提供实时更新的数据和高度个性化的图表定制功能。

关键字 空气质量指数, 数据可视化, 爬虫, Vue, SpringBoot, ECharts

Design and Implementation of Air Quality Information Visualization System based on Crawler

Wu Weiran PengBo Ou Bijie

School of Artificial Intelligence
Nanning University
Nanning 530200, China;

1017035867@qq.com; bopeng8@iflytek.com; bjou@iflytek.com

Abstract—This paper fits the requirements of air quality information visualization system from various angles, and develops a more intuitive domestic air quality information visualization system. The system is composed of real-time AQI data chart display module, historical AQI data chart customization module and data acquisition module. It adopts the technical model of front-end and back-end separation. The front-end framework uses Vue and the back-end framework uses SpringBoot. The air quality data acquisition of the system is obtained by using Python crawler technology, and the chart visualization display of air quality pollution number is realized through ECharts, which the real-time updated data and highly personalized chart customization function are provided.

Key words—AQI; Data visualization; Crawler; Vue; SpringBoot; ECharts

1 引言

随着经济的快速发展和城市化建设的加快, 大气污染状况日益严重, 在人民的生活造成了不良影响^[1]。我国大气污染的主要污染物为颗粒物(PM_{2.5})、颗粒物(PM₁₀), 这两种微粒不仅会对人体的呼吸系统产生危害, 严重时还会威胁到生命。随着工业化和城镇化进程的不断深化, 能源消耗的不断增长, 空气治理的压力也在不断增大。

空气污染的成因有多种。首先是燃料的燃烧, 根据数据显示, 在空气中排放的污染物中, 约有70%~80%是来自于燃煤的北方城市。二是由于工业生产所产生的污染类型比较复杂, 含有多种有毒气体, 所以处理

起来比较困难。三是由于城市化的推进, 各地都在不断扩大基础设施, 土方开挖、场地平整、建筑物拆除、车辆运输等方面的扬尘也随之增加, 成为城市大气污染的重要源头。四是由于汽车排放的影响, 人们的生活水平越来越高, 燃油车的数量越来越多, 而目前国内的汽车依然是以燃油为主, 而汽油中的氢氧化物、颗粒物、挥发性有机物等气体污染物, 也会对城市的空气造成一定的危害^[1]。

2013年以来, 我国出台了一系列大气污染防治法规、政策、标准、行动计划和攻坚方案, 特别是在京津冀及周边地区, 大气污染防治力度前所未有^[3]。环境政策可分为命令控制型、经济激励型、公众参与型, 其中公众参与型政策包括披露环境信息和自愿协议^[4]。

可见民众在治理中起到关键性作用。为此本项目开发一套更加直观的国内空气质量可视化系统,帮助民众了解大气污染治理等相关知识,加强民众的防范意识,全民共同治理,让生活环境更加美好。

2 系统相关技术

2.1 网络爬虫

网络爬虫是一种互联网机器人,用于自动从互联网上获取信息或数据。它实际是一个按照一定的规则从万维网上抓取网页信息的程序或者脚本。机器人扫描并抓取每个所需页面上的某些信息^[5]。实现网络爬虫有两种途径,第一种是使用爬虫软件,第二种是编程语言编写成爬虫脚本。实现爬虫的计算机语言有很多,常用的有 Java、Python、PHP、C++等,本系统主要是使用 Python 来编写脚本。

2.2 Vue 框架

Vue 是一套用于构建用户界面的渐进式框架。它提供了 MVVM 数据绑定和一个可组合的组件系统,具有简单、灵活的 API^[6]。组件是 Vue 最强大的功能之一。组件可以扩展 HTML 页面,封装成可重用的代码。在 Vue 中,父子组件通过 props 参数传递数据,从而实现父组件向子组件单向通讯。在子组件与父组件通讯中,使用监听事件通知父组件改变数据。在开发中遇到组件、HTML 和 JS 等紧密联系时,可以根据开发需求封装自定义组件,能大量减少重复代码编写,提高开发效率。

2.3 Element UI 框架

Element UI 是一款为 Vue 而生的框架,Element UI 遵循用户习惯的语言和行为概念,组件简洁大方,结构和元素保持一致,组件库完备,组件复用性强,能大大提高开发效率。

2.4 ECharts 可视化库

ECharts 是商业级数据图表,具有提供漂亮的图形界面、使用简单、种类繁多、兼容性好等优点,可以流畅地运行在 PC 端和移动设备上,同时 ECharts 可以实现异步数据更新,在初始化图表后可以通过 JQuery 工具异步获取数据再通过 setOption 填入数据和配置相关项即可,其中 setOption 使用 JSON 数据格式^[7]。ECharts 提供了常规的折线图、柱状图、散点图、饼图、K 线图,以及用于统计的盒形图、用于地理数据可视化的地图、热力图、线图、用于关系数据可视化的关系图、旭日图、多维数据可视化的平行坐标、用于 BI 的漏斗图和仪表盘。它还支持图与图之间的混搭^[8]。

2.5 SpringBoot 框架

SpringBoot 是由 Pivotal 开发的全新框架,其设计初衷是用来简化 Spring 开发的。SpringBoot 是基于 Spring4.0 设计出来的,不仅继承了 Spring 框架原有的优秀特性,而且还通过简化配置来进一步优化了 Spring 应用的整个搭建和开发过程。另外, SpringBoot 通过集成大量的框架使得依赖包的版本冲突,以及引用的不稳定性等问题得到了很好的解决^[9]。SpringBoot 具有以下优点:

(1) 开发人员可借助 SpringBoot 高效的搭建好一个 Spring 项目,不需要像以往创建一个 Spring 项目那样,要引入大量的依赖包,配置大量的 XML 文件,就能集成其他框架,大大缩短了项目的搭建时间,让开发者的重心放在业务逻辑的编写上。

(2) 以往的 web 应用需要部署在服务器上时,若服务器没有安装 Tomcat 环境,需要先安装环境才能部署。而 SpringBoot 内置了部分 Servlet 容器,不需要打包成 WAR 包,能直接将项目打包成 JAR 包,直接在项目中依赖 Java 环境运行。

(3) starter 自动依赖与版本控制。starter 启动器的自动依赖功能帮助开发者导入项目中的全部依赖,并自动进行版本控制^[10]。

3 系统的概要设计

3.1 总体功能设计

空气质量可视化平台主要由实时 AQI 数据图表显示模块、历史 AQI 数据图表自定义模块、数据获取模块组成。系统的整体业务流程主要是对爬虫获取的数据进行整理,然后通过 ECharts 进行图表展示。总体功能设计如图 1 所示。

3.2 总体架构设计

考虑到了系统开发的可维护性、成本和拓展性,空气质量可视化平台采用完全前后端分离的技术来设计,系统的整体架构主要包括访问层、表示层、交互层、业务逻辑层、数据层。系统架构如图 2 所示。

用户通过访问层访问系统平台,表示层通过 Vue 的路由渲染出页面元素与用户进行交互,表示层通过 Element UI 美化页面布局,并搭配主流的可视化框架 ECharts 进行图表渲染。

表示层通过交互层向业务逻辑层发送 axios 异步的数据请求,业务逻辑层对用户行为产生的请求进行处理,响应成功后将响应结果返回到表示层,进行页面渲染^[11]。业务逻辑层主要通过 SpringBoot 技术实现,Controller 层监听请求的接口,对参数进行解析后利用依赖注入技术进行业务逻辑处理,最后通过 mapper 层与数据层 MySQL 数据库进行交互,并将请求结果

返回。

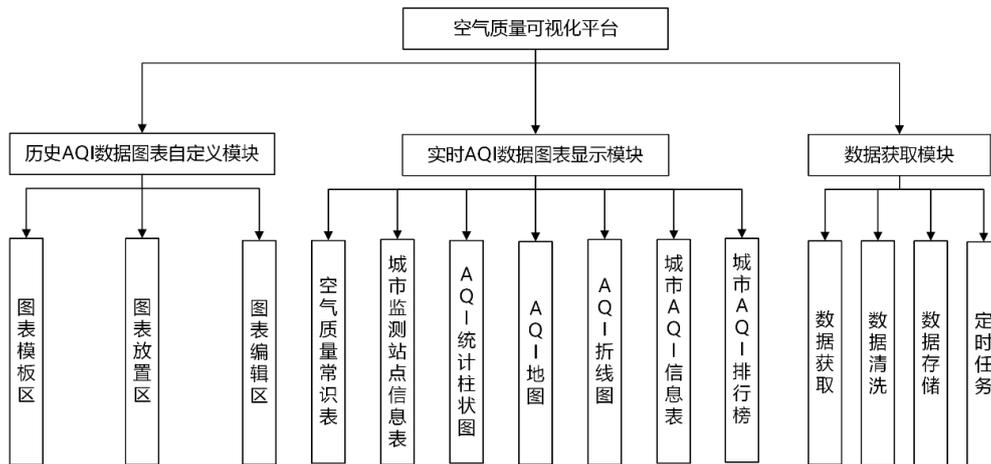


图 1 系统总体功能设计图

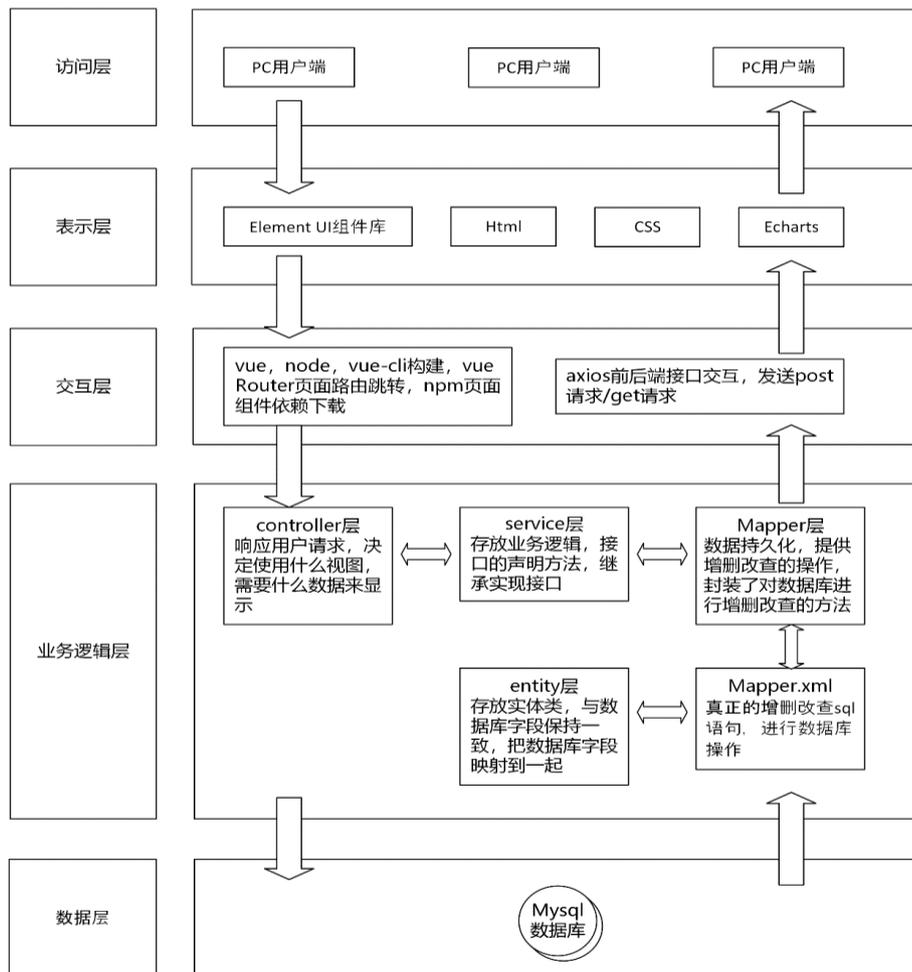


图 2 系统架构设计图

3.3 数据库设计

分析了需求的各个模块后，本系统所涉及的实体类、属性和实体类之间的关系如图 3 的 E-R 图所示。

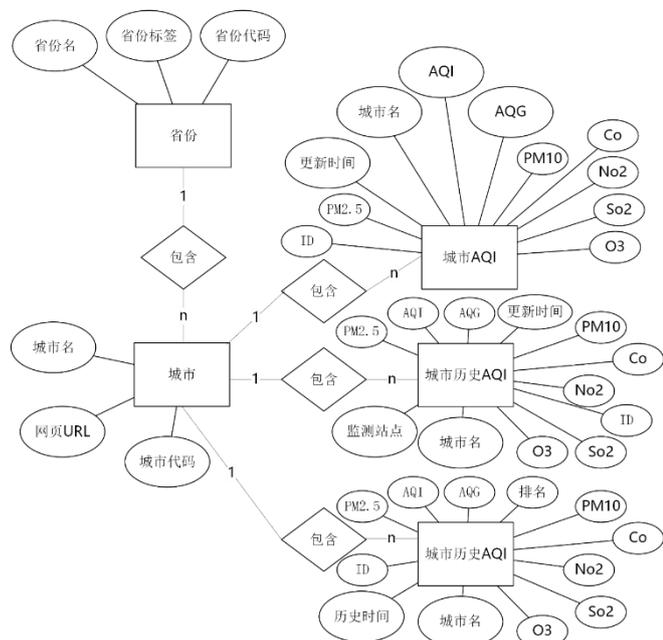


图 3 空气质量可视化系统 E-R 图

4 系统详细设计

(1) 数据获取模块设计

该模块主要是由爬虫程序组成。根据系统功能需求来挑选爬取网站，挑选条件是网站必须包含详细的 AQI 数据、网站 AQI 数据是实时更新的、网站保留了历史 AQI 数据。爬虫程序的流程图如图 4 所示。

由图 4 可知，数据获取模块整个业务流程是由启动定时器开始的，定时器启动后，开始判断是否符合爬取时间，若未达到符合时间则定时器继续计时，若时间符合，则向数据库获取网址生成 URL 队列，对网页信息进行抓取，抓取完成后进入异步操作，程序分成两步，一步是程序判断是否关闭计时器，若不关闭则定时器继续计时，若关闭则程序结束；二是解析数据，将解析完成的数据存入数据库，并生成日志文件存储至本地。

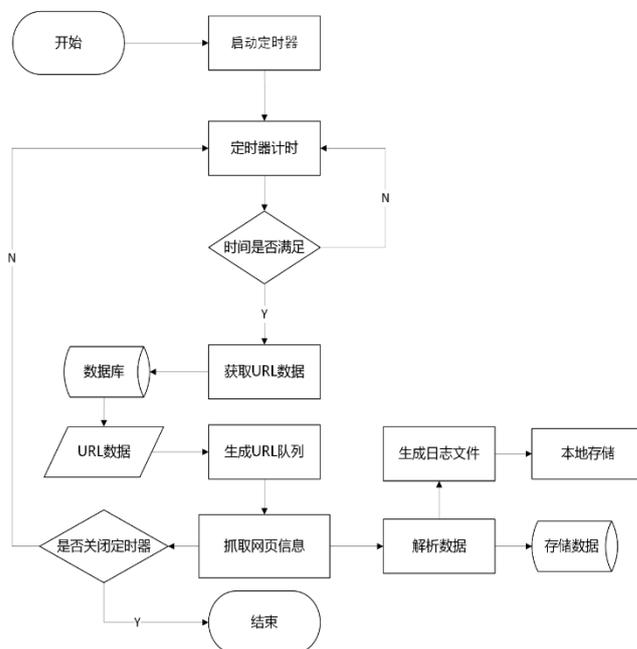


图 4 数据获取模块流程图

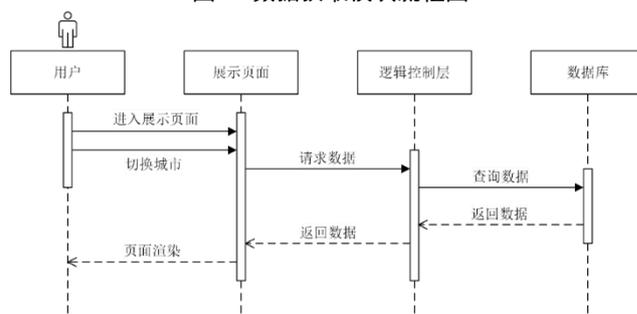


图 5 实时 AQI 数据图表显示时序图

(2) 实时 AQI 数据图表显示模块设计

该模块是系统的主要展示区，由系统自动创建。用户可以根据需要与图表进行交互。显示模块时序图如图 5，活动图如图 6。

(3) 历史 AQI 数据图表自定义模块设计

该模块可由用户对历史 AQI 数据进行自定义图表生成。用户可以根据需求选择图表模板、配置图表数据值、图表的样式等。

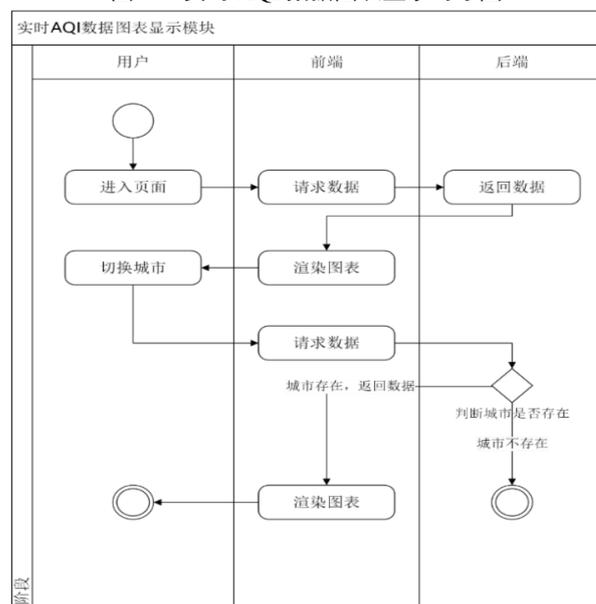


图 6 实时 AQI 数据图表显示活动图

结合子模块的三个区域可知，图表自定义模块活动图如图 7 所示。

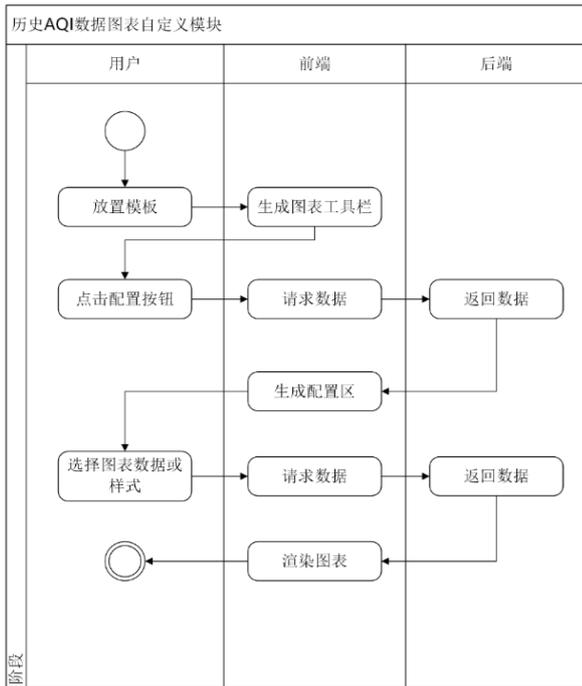


图 7 历史 AQI 数据图表自定义模块活动图

5 系统的实现

(1) 数据获取

在本文中提出，本系统的数据获取依靠爬虫程序，所以选择合适的网站是本系统开发的首要阶段。在对网上提供空气质量指数的网站进行筛选后，最终选择天气后报网作为本系统的数据来源。天气后报网的网址为 <http://www.tianqihoubao.com/aqi/>。

```

抓取数据中...
['上海', '普陀', '36', '优', '24', '9', '1.7', '11', '2', '113', '2022-04-30 22:00']
['上海', '十五厂', '33', '优', '0', '16', '0.9', '12', '5', '104', '2022-04-30 22:00']
['上海', '徐汇上师大', '35', '优', '30', '24', '1', '16', '8', '94', '2022-04-30 22:00']
['上海', '杨浦四漂', '0', '-', '0', '15', '0.5', '10', '8', '99', '2022-04-30 22:00']
['上海', '静安监测站', '33', '优', '33', '10', '0', '15', '4', '102', '2022-04-30 22:00']
['上海', '浦东川沙', '33', '优', '20', '12', '0.9', '15', '4', '104', '2022-04-30 22:00']
['上海', '浦东新区监测站', '33', '优', '24', '15', '0.7', '15', '5', '103', '2022-04-30 22:00']

```

图 8 获取城市监测点空气质量指数信息实现效果图

(2) 实时数据可视化模块实现

实时 AQI 数据图表显示模块是本系统的主页面，是可视化平台的主要展示区，页面如图 9 所示。

(3) 历史数据可视化模块实现

历史 AQI 数据图表自定义模块的主要功能是，用户可以使用丰富的图表模板对城市历史 AQI 数据进行图表自定义生成，用户可在自定义过程中，随时切换新的数据值和图表样式，此外图表还可以下载和全屏展示。该模块的主页面如图 10 所示。

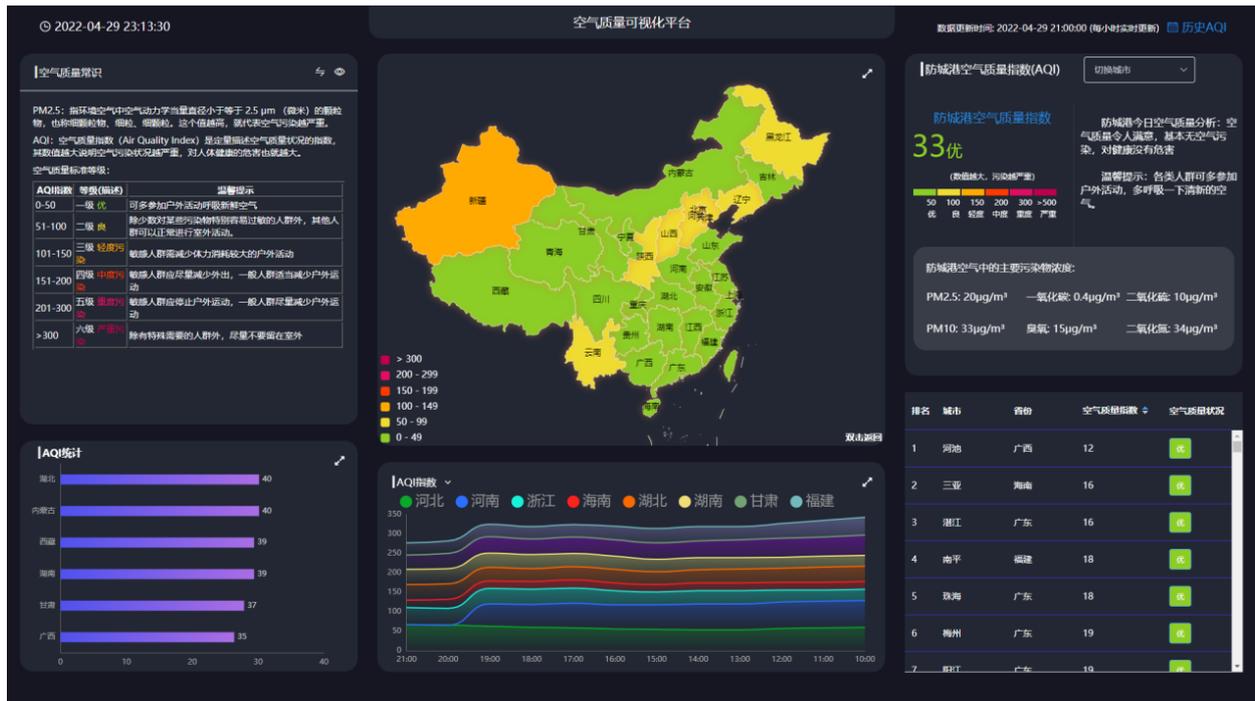


图 9 实时 AQI 数据图表展示区

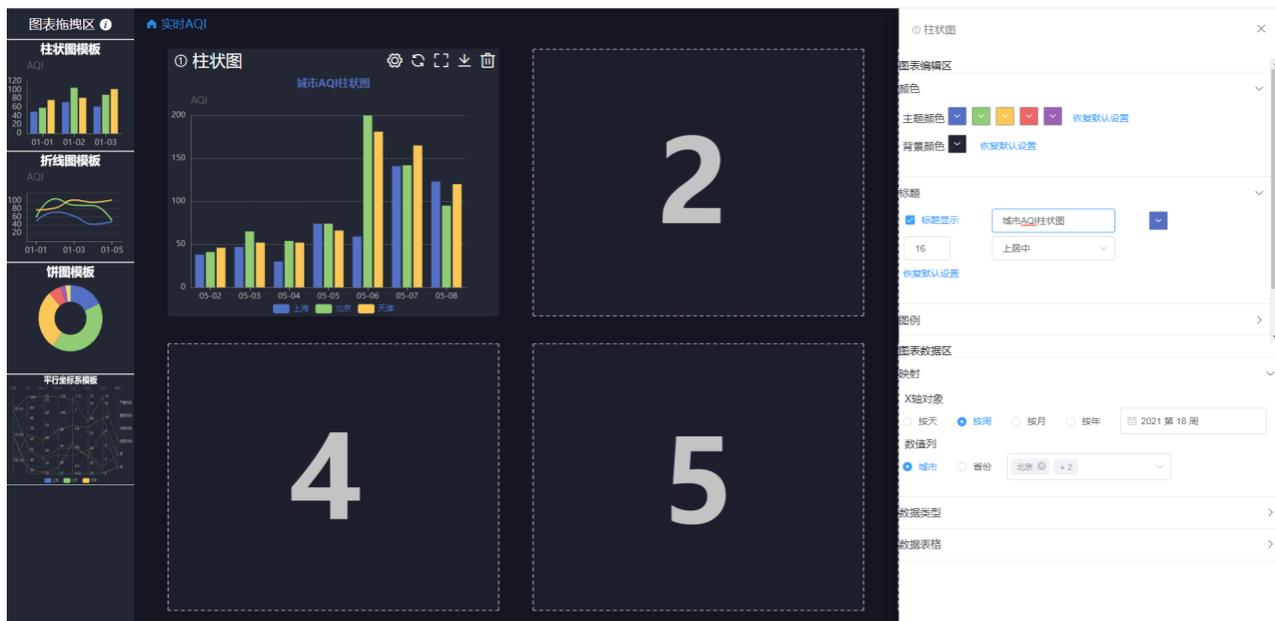


图 10 历史 AQI 数据图表自定义模块

6 结束语

本文主要叙述了基于爬虫空气质量信息可视化的设计与实现，介绍了研究背景、系统相关技术、系统需求分析、系统详细设计和系统模块详细设计等内容。空气质量信息可视化系统拥有实时数据展示，能够对用户的出行提供较大的帮助。自定义图表模块能让用户也能够参与数据呈现这一过程，让用户在使用中学习空气质量的的相关知识，提高此类环境污染问题的关注度，进而提高居民的防范意识，为大气污染治理事业献出一份薄弱的贡献。

参考文献

- [1] Zaib S, Lu J, Bilal M. Spatio-Temporal Characteristics of Air Quality Index (AQI) over Northwest China. 2022.周宏春.大气污染防治行动计划[J].绿色中国,2017(08):54-56.
- [2] 田园.大气污染的成因及治理措施分析[J].清洗世界,2021,37(12):70-71.
- [3] 冯贵霞.大气污染防治政策变迁与解释框架构建——基于政策网络的视角[J].中国行政管理,2014(09):16-20+80.
- [4] 姜华,高健,李红,储王辉,柴发合.我国大气复合污染协同防控理论体系框架初探[J/OL].环境科学研究:1-15[2022-01-15]
- [5] 冯艳茹.基于 Python 的网络爬虫系统的设计与实现[J].电脑与信息技术,2021,29(06):47-50
- [6] 柴青山.基于 MVVM 模式的 Vue.js 框架在物流软件自动化测试系统中的应用研究[D].北京:北京邮电大学,2019.
- [7] 袁江琛.基于 jQuery+Ajax 的 ECharts 在成绩统计中的实现[J].智能计算机与应用,2019,9(05):217-219+222.
- [8] 王志文.Vue+Elementui+Echarts 在项目管理平台中的应用[J].山西科技,2020,35(06):45-47.
- [9] 邓笑.基于 Spring Boot 的校园轻博客系统的设计与实现[D].武汉:华中科技大学,2018.
- [10] 张帅.智慧城市大数据可视化云平台的设计与实现[D].沈阳:沈阳大学,2021
- [11] 贺紫璐.基于 SpringBoot 和 Vue 框架的第三方医疗器械供应链平台的设计与实现[D].衡阳:东华大学,2019