

作品设计书

——云上政声：政府动态信息分析与可视化系统

目录

- 一. 作品设计背景和意义1
- 二. 现状及内容1
- 三. 需求分析2
 - 3.1 用户需求分析：2
 - 3.2 可行性分析：2
- 四. 系统功能实现2
- 五. 开发技术3
- 六. 框架设计3
- 七. 界面设计与数据库设计4
 - 7.1 界面设计：4
 - 7.2 数据库设计：4
- 八. 三个模块的主要内容介绍：4
 - 8.1 spider 爬虫模块4
 - 8.2 database 数据库模块5
 - 8.3 Vue 交互模块5
- 九. 设计模块流程图：5
- 十. 前后端架构：6
 - 10.1 后端程序6
 - 10.2 前端程序7

一. 作品设计背景和意义

随着信息技术的发展和社会信息化的加速推进，公众对政府信息公开和透明度的要求越来越高。政府网站作为政府信息公开和透明度的重要平台之一，如何通过技术手段提高政府网站信息的获取和利用效率，已经成为当前政府工作和社会发展的课题。随着数据挖掘和分析技术的不断发展和完善，如何通过数据挖掘和分析技术挖掘和分析政府网站信息的重要性越来越被人们所重视。如何利用数据挖掘和分析技术对政府网站信息进行深入分析和挖掘，以便更好地服务于公众和政府的决策，也成为当前技术研究和实践的热点。

作为现代社会管理和公共服务的重要手段，政府网站不仅是政府向公众提供信息和服务的重要渠道，也是政府与公众之间交流和互动的重要平台。如何通过技术手段提高政府网站的信息传递和交流效率，以便更好地服务于公众和社会管理，也成为当前社会管理和公共服务的重要需求。

通过自动化获取政府网站信息并进行数据分析和可视化展示，促进政府工作的透明度和公开性，让公众更加了解政府工作的真实情况和决策过程，增强政府工作的公信力和可信度。通过更加直观和易懂的形式展现政府信息，促进政府与公众之间的良性互动和合作关系，增强公众对政府工作的了解和参与感，促进政府与公众之间的沟通和理解。通过提供数据分析和可视化结果，让公众更加全面地了解政府工作的情况和政策导向，增强公众对政府的信任和支持，促进社会稳定和谐发展。此外，政府网站作为政府向公众提供服务的重要渠道，通过自动化获取政府网站信息并进行数据分析和可视化展示，还能够更加精准地了解公众需求和诉求，更好地为公众提供优质的公共服务和社会管理。因此，该项目的研究和实践具有重要的现实意义和社会价值，有助于推动政府信息化和社会管理现代化的进程，促进社会的和谐稳定发展。

根据以上设计依据和意义，本项目设计了一款基于前后端分离的政府动态信息分析与可视化系统。整个项目的实现过程分为数据采集、数据存储、数据处理、数据展示四个步骤，通过上述技术实现，可以完成对政府网站信息的自动化采集、存储、处理和展示，从而实现对政府工作的监督和公众参与的促进。

二. 现状及内容

在政府信息化领域，已有不少学者和专家通过技术手段，对政府网站信息进行研究和分析，以实现政府工作的透明、公开和规范化。其中，数据挖掘、文本分析、机器学习等技术是较为常见的研究手段。在数据挖掘方面，研究者通过挖掘政府网站的数据，分析政府工作的趋势和热点问题，为政府制定决策提供参考。例如，利用数据挖掘技术对政府官方微博的数据进行分析，可以发现政府官方微博在传递政策信息、促进社会互动等方面的作用和影响。在文本分析方面，研究者通过对政府文件、公告等文本的分析和处理，探究政府工作的规范和效率。例如，利用文本分析技术对政府文件的语义结构进行分析，可以发现政府文件中存在的问题和矛盾，提出优化建议。在机器学习方面，研究者通过对政府网站的数据进行训练，实现对政府信息的自动化处理和管理。例如，利用机器学习技术对政府网站的数据进行分类和标记，可以实现对政府工作的快速检索和管理。在本项目中，研究者将结合数据采集、存储、处理和展示等多种技术手段，对政府网

站的信息进行全面的分析和研究,实现对政府工作的全面监督和公众参与的促进。

三. 需求分析

3.1 用户需求分析:

此系统主要是基于具有较强公信力的贵州省人民政府网站和贵州省投资促进局。在该项目中,用户主要包括政府相关部门和广大民众,对于不同的用户,需求有所不同,以下是该项目的用户需求分析:

1. 政府部门: 政府部门作为信息发布方,需要通过该项目了解民众对政府工作的关注点和热点问题,以便更好地调整和优化政策和工作计划。

2. 研究人员: 研究人员需要通过该项目获取相关政策和热点问题的数据,以便分析和研究政府工作的趋势和变化。

3. 媒体人员: 媒体人员需要通过该项目了解政府工作的最新动态和热点问题,以便及时报道和传递信息。

4. 普通民众: 普通民众需要通过该项目了解政府工作的最新动态和政策,以便更好地参与社会生活和维权。

在用户需求分析的基础上,项目设计应该注重数据的准确性和完整性,确保数据能够真实地反映政府工作的实际情况。同时,该设计还应该注重数据的可视化和易用性,以使用户能够方便地浏览和理解数据。最后,该项目还应该注重用户反馈和改进,不断完善和优化项目的功能和体验。

3.2 可行性分析:

从开发工具和数据存储方面来看,本次设计所需的开发工具是 Pycharm、IDEA、VSCode 和 Navicat,学生只需用学校邮箱即可申请免费使用。启动前后端程序之后,只需在浏览器地址栏输入 localhost:5174 即可使用系统

从技术方面考虑,现在 Python 作为最强大的编程语言之一,拥有很多功能强大的第三方生态,在数据爬取方面采用 requests,数据解析采用 lxml,数据库连接采用 pymysql。后端使用 java 生态下的 springboot 技术,通过 mybatis 连接数据库,使用 jieba-java 做数据分析。前端使用 vue 框架,通过 axios 请求后端服务获取 json 数据,经过 echarts 可视化图表

本系统功能强大,使用简单。可以在市州动态页面指定日期和时间快速了解一个地级行政区的发展方向和工作重点。并且通过数据大屏页面可以查看整个省政府的关注领域和地区发展等情况。

四. 系统功能实现

本设计包含四项功能,分别是数据采集功能,数据存储功能,数据展示功能,数据分析功能,以下展示了各个功能的作用:

数据采集功能: 通过网络爬虫等技术手段,自动化地从政府网站的地方动态专栏中抓取每一个词条的相关信息。

数据存储功能: 将爬取的数据通过第三方库 pymysql 连接到 MySQL,存入到 gov 数据库下 gov 和 location 两张表。

数据展示功能: 通过 BS 架构将服务器与浏览器连接,在服务器将数据处理后,把结果以图表等形式直观的在浏览器呈现给用户,方便用户查看和分析。

数据分析功能：实现对每个词条中的文本内容进行分词、统计词频等分析。

五. 开发技术

本次项目设计涉及到以下几个方面的技术：

1. Web 爬虫：用于从省政府网站的地方动态专栏下抓取每一个词条，可以使用 Python 的 Requests 和 lxml 等库实现。

2. 数据库管理系统：用于将抓取的词条保存到 Mysql 数据库中，可以使用 Python 的 pymysql 等库连接和管理 Mysql 数据库。

3. 数据分析和可视化：用于对抓取的词条进行词频统计，生成柱状图和饼图，并通过 Web 视图回显给用户，直观的展示词频情况。

总体来说，这个项目主要采用 Python 和 Java 作为开发语言，并利用生态系统中丰富的第三方库和工具来实现各个功能模块。同时，还需要对 Vue 框架有一定的了解，以便实现良好的用户交互界面。

本项目以贵州省为例，从贵州省人民政府市县动态以及贵州省投资促进局地方动态爬取所有地方动态的数据，并将数据保存到 MySQL 里面。通过 Web 浏览器视图在主页由用户自主选择时间段，程序将从数据库里检索出相应的记录，并返回到前端界面上，还会统计取出来的数据里，哪些市州（通过 jieba 分词）出现的频次比较多，以此为依据生成图表。

六. 框架设计

该项目的框架设计为 B/S 架构，B/S 架构即浏览器和服务器架构模式，是随着 Internet 技术的兴起。在这种架构下，用户工作界面是通过 Web 浏览器来实现，极少部分事务逻辑在前端（Browser）实现，但是主要事务逻辑在服务器端（Server）实现，形成所谓三层 3-tier 结构。

B/S 架构是 WEB 兴起后的一种网络架构模式，WEB 浏览器是客户端最主要的应用软件。这种模式统一了客户端，将系统功能实现的核心部分集中到服务器上，简化了系统的开发、维护和使用。本项目可以考虑采用以下几个模块：

1. 数据爬取模块：负责从贵州省政府网站的地方动态专栏下抓取每一个词条，并将其保存到 Mysql 数据库中。

2. 数据分析和可视化模块：负责对抓取的词条进行词频统计，并通过浏览器反馈显示在给用户。

3. 数据库管理模块：负责连接和管理 Mysql 数据库，并提供与数据爬取模块和数据分析和可视化模块的交互接口。

4. 数据展示模块：在主页给出了用户的注册和登录界面，同时可由用户自主选择时间段，并分别展示了不同市的各时间段的不同词汇频率，生成图表进行直观的展示。

在这个框架设计中，各个模块之间通过接口进行交互，实现了各自的职责划分，提高了代码的可维护性和可扩展性。同时，通过将数据爬取和数据分析和可视化模块分离，可以更好地实现数据的处理和展示功能。

七. 界面设计与数据库设计

7.1 界面设计:

在界面设计中,应该注重用户交互的友好性和易用性,提供简洁明了的操作界面和丰富的可视化效果,让用户能够方便快捷地使用该系统,并获得实时的数据反馈和展示效果。因此,大致有以下控件:

1. 时间段的选择控件和各市区下拉选择菜单;
2. 用户的注册和登录界面
3. 用户的个人设置界面
4. 市州区县的分时间段和地区的图表展示,和关键词的图表展示,分为柱状图和饼图。
5. 数据大屏:贵州省的整体数据集成展示图,包含了行业分布,时间分别,领域分布,和一些城市情况。

7.2 数据库设计:

数据库设计是两张表,一张用于存储贵州省各级地区的名字,另一张表是 id, title, href, data, ref, 分别代表主键, 词条标题, 词条超链接, 词条日期, 词条所属地区。这种设计可以很好地存储并查询词条的相关信息,同时也有助于后续的统计和分析。

地区名字表: 该表用于存储贵州省各级地区的名字,可命名为 locations。该表主要包含两个字段:地区 id 和地区名称。其中,地区 id 可以设置为自增主键,地区名称可以设置为唯一索引,以避免重复存储相同的地区名称。

词条信息表: 该表用于存储地方动态专栏下的词条信息,可命名为 gov。该表主要包含以下字段:

- id: 自增主键,用于唯一标识每条词条信息。
- title: 存储词条标题,类型为 VARCHAR 或 TEXT。
- href: 存储词条的超链接,类型为 VARCHAR 或 TEXT。
- date: 存储词条的发布日期,类型为 DATE 或 DATETIME。
- Ref: 外键,与地区名字表中的地区 id 相关联,用于表示该词条所属地区。

通过这样的数据库设计,可以很方便地存储和查询地方动态专栏下的词条信息,并通过 ref 字段与地区名字表进行关联,实现词条所属地区的查询和统计等功能。

八. 三个模块的主要内容介绍:

8.1 spider 爬虫模块

在该模块下,主要功能由核心类 GovSpider 完成,有以下方法:

- get_gov_name: 爬取贵州省所有的市州区县
- spider_gov: 爬取贵州省人民政府网的地方动态
- spider_invest: 爬取贵州省投资促进局的地方动态

同时可以在实例化 GovSpider 的时候,传入指定的 url 和 xpath 规则,调用

上述方法依然能爬取到指定的数据

除开市州区县的数据外，省政府以及投资促进局的数据都是一个二维列表，每一个元素包含标题，URL 链接以及日期。

总之，该模块会通过指定的 URL 发送 HTTP 请求获取网页的 HTML 内容，然后使用解析库从 HTML 中提取出目标数据，并存储到数据库中

8.2 database 数据库模块

在该模块下，主要功能由 Mybatis 和 Springboot 完成，提供了以下接口：

- /login: 负责用户登录
- /getOneYearNews: 负责返回最近一年内所有地区的数据
- /getSpecifyDateNew: 返回指定日期端的词条数据
- /getSpecifyDateAndRegionNews: 返回指定时间段和地区的词条数据
- /getSpecifyDateAndRegionTop: 返回指定时间段和地区的关键词

综上，该模块主要负责数据的存储和管理，包括数据表的创建、数据的插入、更新、查询和删除等操作。该模块使用 MySQL 数据库，通过 Java 编程语言连接 mybatis 来实现对数据库的操作。

8.3 Vue 交互模块

该模块提供了三个部分，第一个部分提供了用户的注册和登录，当用户登录进入之后进入第二部分，第二部分是整体内容展示，里面分为三块，第一块的个人中心的信息和修改，第二块是各市州区县的内容展示，第三块是数据大屏，也就是数据的整体展示。

九. 设计模块流程图：

见图 9.1

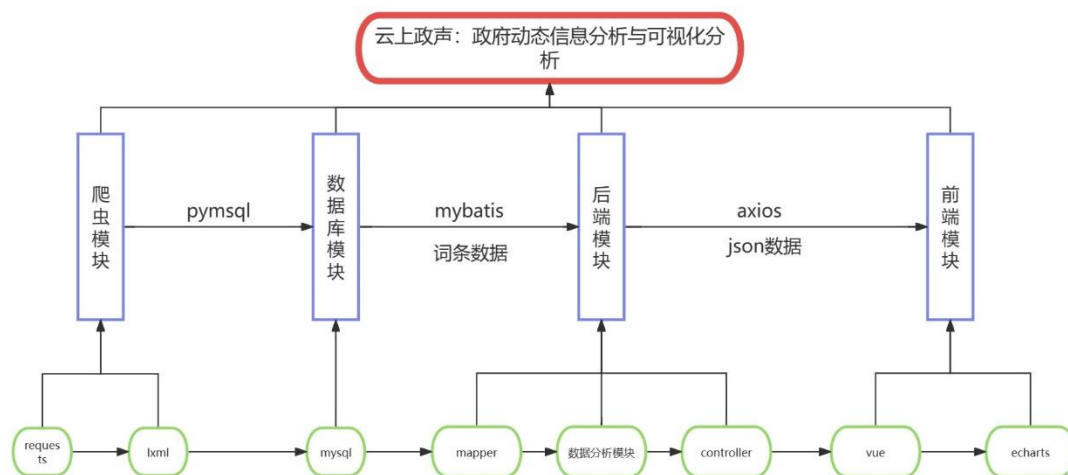


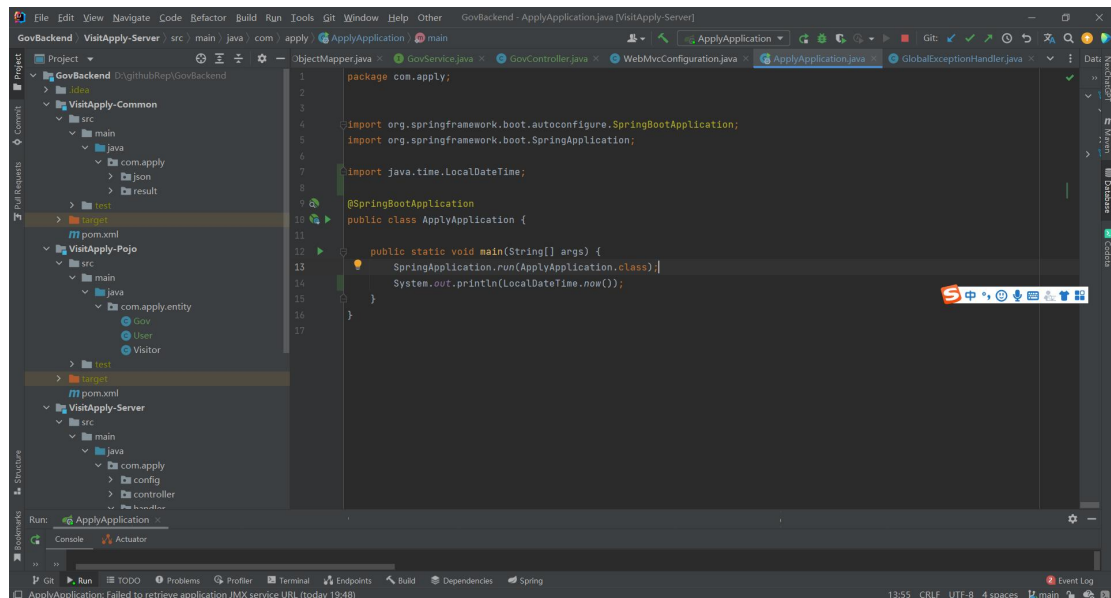
图 9.1 系统模块详解图

十. 前后端架构:

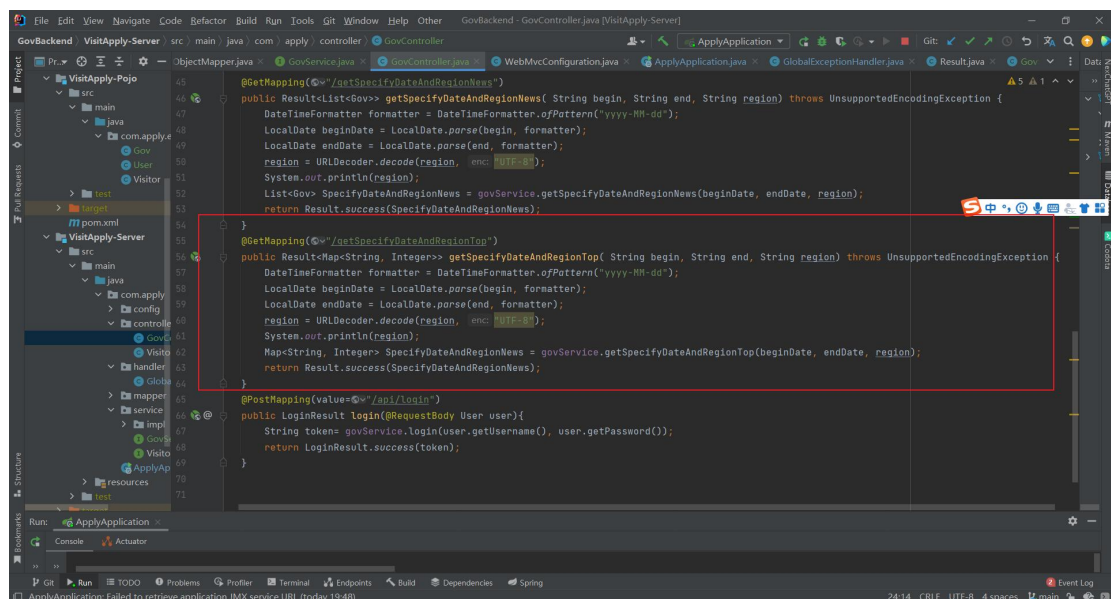
10.1 后端程序

后端使用 springboot、maven、mybatis、mysql 等技术。使用 maven 聚合成三个模块:

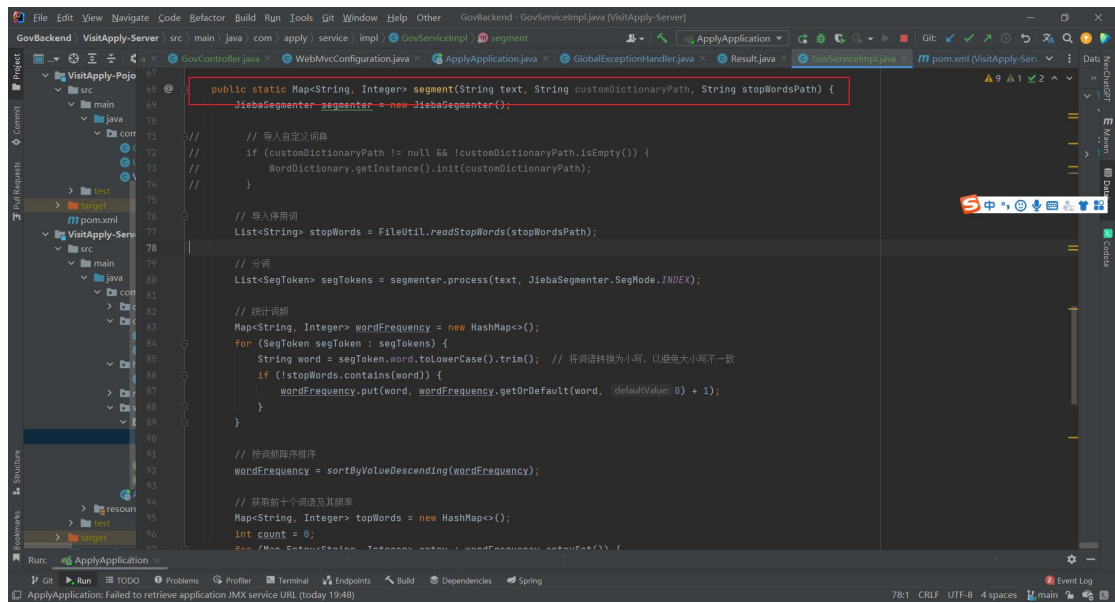
- 1.Common: 定义 json 序列化相关类, 以及统一相应结果类
- 2.Pojo: 定义与数据库表一一对应的实体类
- 3.Server: 后端服务的核心, 有 controller、service、mapper 以及全局异常处理器



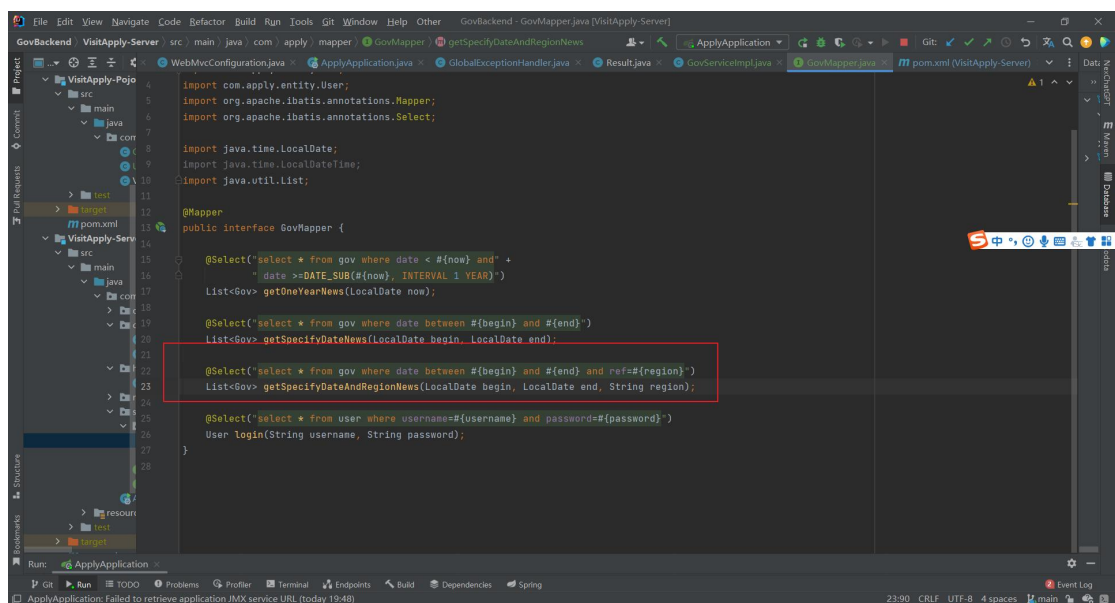
后端最重要的一个接口是/getSpecifyDateAndRegionTop, 这个参数接收字符串形式的起始时间、截止时间、所属区域, controller 将参数格式化交给 service



service 中核心方法是 getSpecifyDateAndRegionTop, 其中调用了自定义私有方法 segment, 这个方法是系统负责数据分析的模块, 包括导入自定义词典、导入停用词、分词、统计词频、按词频降序排序、获取前十个词语及频率等功能



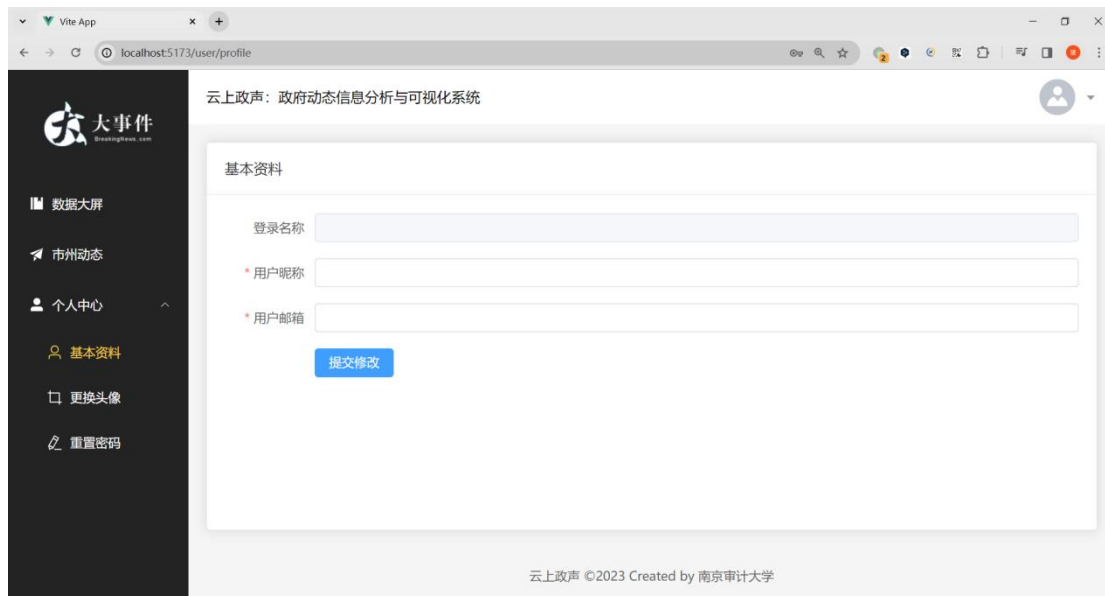
service 中 mapper 中 getSpecifyDateAndRegionNews 方法会从数据库中检索出符合日期和地区条件的数据



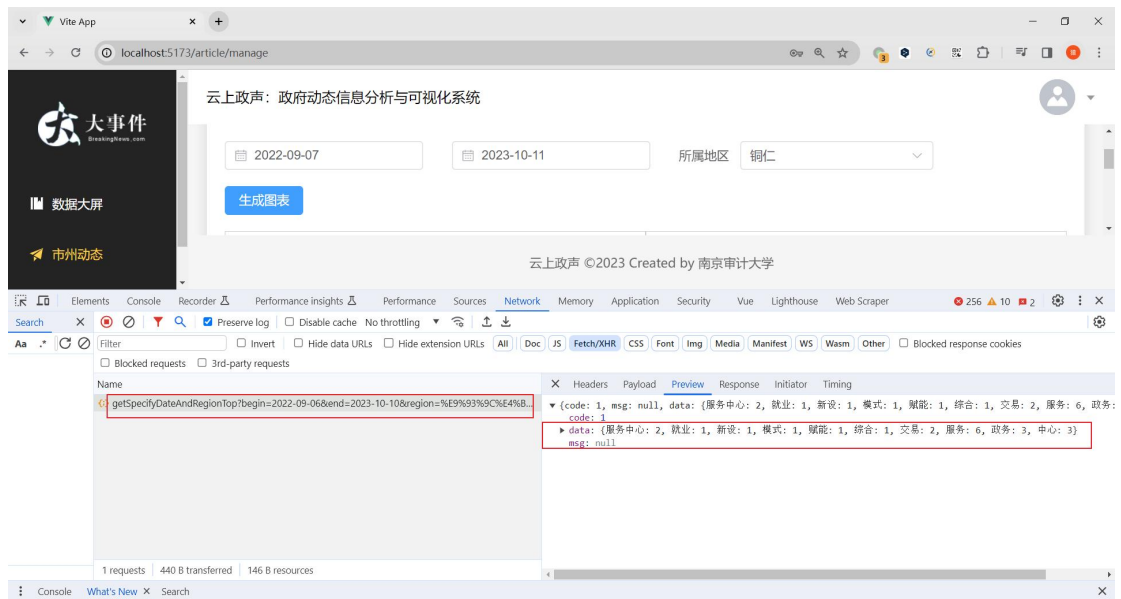
10.2 前端程序

前端使用 Vue3、ElementPlus、Axois、Echarts 等。一共有登录页、市州动态、个人中心、数据大屏四个页面





在市州动态页面，当选定好日期跟地区后，单击生成图表会发起一个 ajax 携带参数请求后端的 `/getSpecifyDateAndRegionTop` 接口服务，得到相应的 json 数据



同时经过 echarts 将后端返回的 json 数据渲染成柱状图和饼图

