

《web 编程》课程实验报告

实验作品名称	基于动态爬虫的个人新闻管理系统		
姓名	郭夏辉	学号	10211900416
电话		电子邮件	10211900416@stu.ecnu.edu.cn

原创承诺

本人承诺本作品为本人独立创作完成，不侵犯任何第三方的知识产权或其他权利，同时符合中华人民共和国的有关法律，如本作品发生侵权行为由本人自负责任。

作品简介

（设计目标与意义、关键技术、作品特色）

掏出你的手机，打开抖音、快手这样的短视频平台已经成为了当今人们主流的获取信息的方式。短视频加速了新闻的传播，丰富了受众的体验，但也有碎片化和真实性偏低的问题。传统的新闻媒体凭借多年来积累的公信力和权威性，依托官方的消息通道，依然是我们获取资讯的重要来源。它们通过升级新闻门户网站，提升了我们读者的阅读体验，响应了数字时代的大趋势。但是面对排山倒海的消息，我们的时间精力是有限的，只能挑选感兴趣的几条来阅览。对于来自上游媒体的热点信息，下游的媒体基本都会报导，这就带来了信息的冗余，所以理论上来说我们只需要选择并获取几个有代表性的新闻网站讯息即可，在此我选择的是新浪新闻、网易新闻、东方财富这三个知名平台。相信今天您就是海量信息流的主人。

对于新闻这样的结构化数据，我使用 MySQL 来存储。新闻有着源源不断产生的特点，不像已有的数据集，需要利用爬虫来搜集起来。通过分析新闻页面的报文结构，我创造性地使用了动态爬虫技术，只需要在前端点击一下就能获取全部的信息，而不是说先独立地把运行完爬虫把数据写入到数据库后才能用前端演示。前端我采用了 Vue 框架，这是一套轻量级的渐进式前端框架，提升了项目的效率、节省了开发的成本，然后我使用 element-ui 组件库、axios 网络请求库及 echarts 可视化插件丰富了功能；后端我采用 express 框架，它不仅性能高，而且提供了丰富的插件和中间件，可以轻松地扩展和定制应用程序的功能。后端主要负责查询数据库中的用户及新闻信息，并给前端提供了增删改查的接口。整体来说前端、后端、爬虫、数据库相互分离，项目耦合度较低，遵循 RESTful 架构的原则。

基于个人新闻管理系统，您不仅可以随心操纵来源于三个平台的海量新闻，而且还能根据新闻的标题、内容等信息进行有效检索，分页浏览使得界面更加简洁。如果您特别关心诸如"某明星"这样的关键词，这个系统还提供了时间热度分析看板，在可视化的词云展示中洞悉最核心的内容。最后，这套系统也为您提供了用户注册、登录、昵称这样的功能，充分保证了您的隐私。如果您对此系统有什么不满意的地方，还能通过意见反馈面板向我指出。整个项目的细节处我也加入了一些美化，比如点击右键有烟花弹幕般的效果，增进了趣味性。

作品开发技术及运行环境

Node.js v22.4.0

前端我采用 vue 框架，使用 element-ui 组件库、axios 网络请求库以及 echarts 可视化插件，详见 vue/package.json

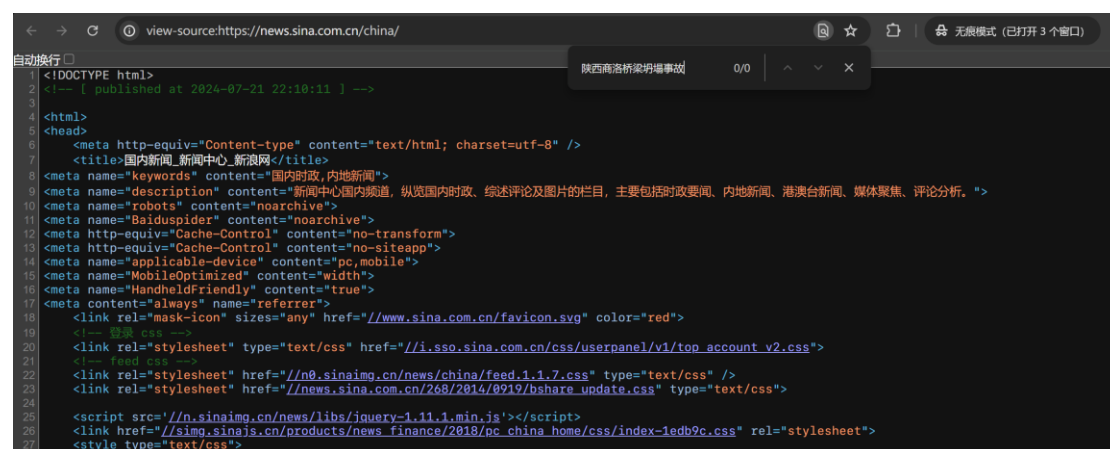
后端我采用了 express 框架，负责操作数据库，管理其中的用户及新闻信息，依赖库详见根目录下的 package.json

实验步骤（具体说明制作的过程）

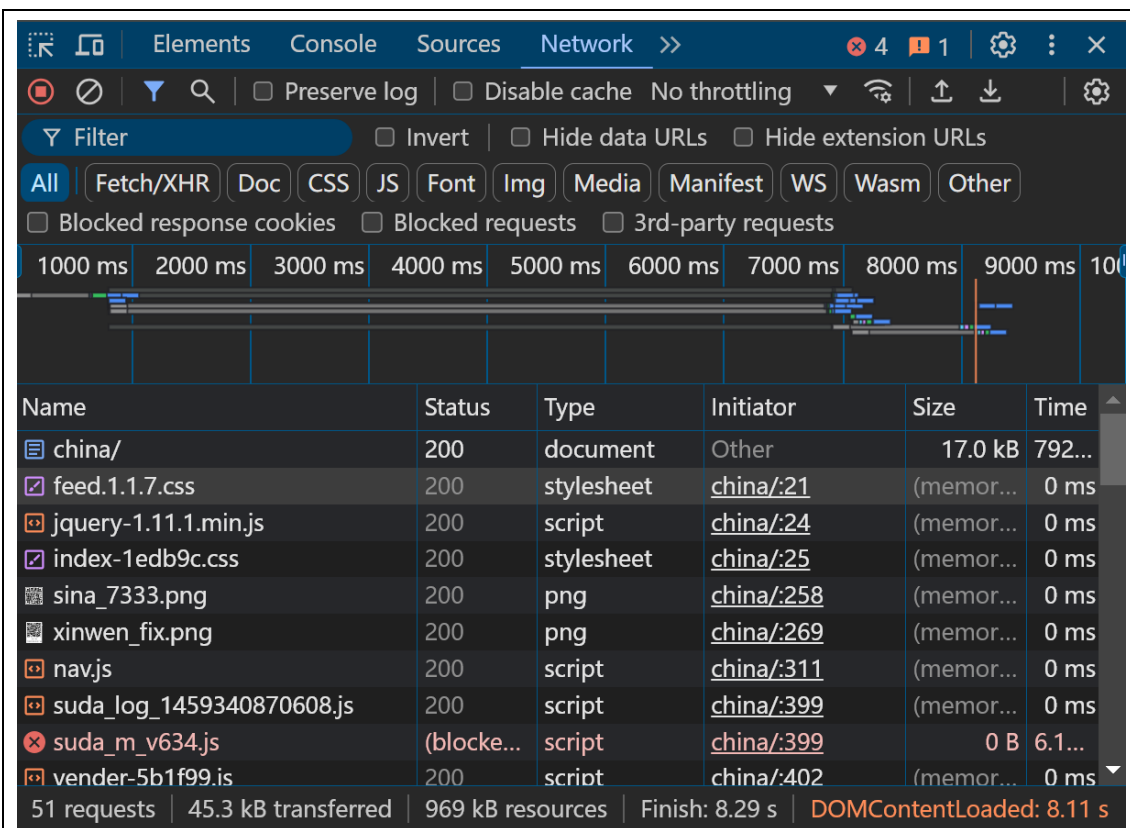
整个实验的过程我主要分为爬虫设计、后端、前端三个板块。因为篇幅有限我不可能面面俱到，只能挑选重要的进行阐述。

一、爬虫设计

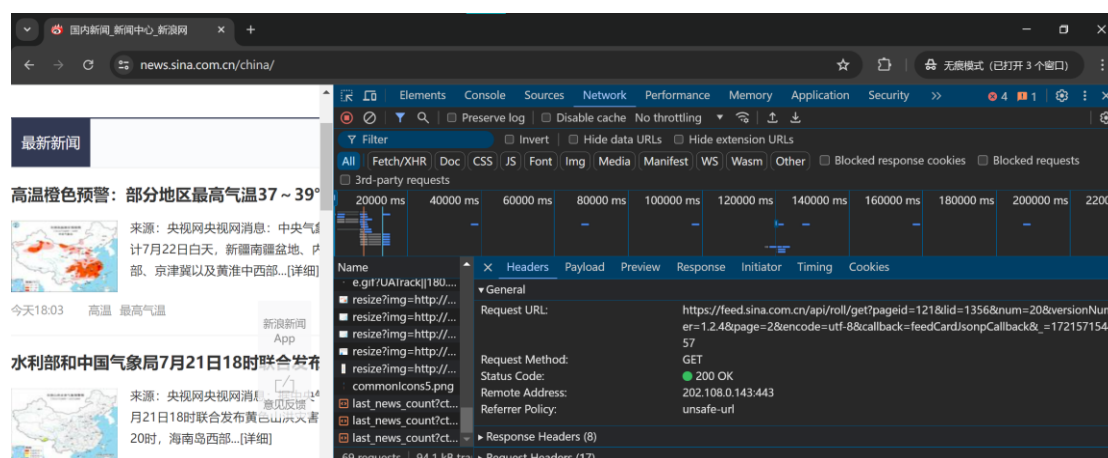
在选择了三个新闻网站之后，我们需要对它们的网站进行分析，以此来确定爬取的方式，这里我以爬取新浪新闻国内板块（<https://news.sina.com.cn/china>）为例。在这个页面按下观察网页源代码，先搜索第一条新闻标题（陕西商洛桥梁坍塌事故）发现根本找不到，说明这是一个动态页面：



接下来怎么办呢？我在 F12 窗口下的 Network 中要找到那条请求，但是请求非常多我怎么找呢？



我用了一个比较巧妙的机制来查找。比如说当前是第一页，把 Network 拉到最底下，再点下一页会有多出来的部分。多出来的就是我们想要的链接，这样的话就可以更高效的找到目标请求：



响应内容长这样：

入并配置 CORS 中间件，最后使用路由模块并启动服务器，监听 3000 端口。在这个过程中，值得注意的是需要 `app.use(cors())` 来允许跨域请求否则便会报错。

每个路由文件都实现了 `restful` 分隔接口，还实现了用户登录、注册、信息修改的接口。具体的实现见 `router` 下面的各个路由文件，简而言之用就是与前端对接，然后通过一系列的 SQL 语句进行增删改查。我考虑到删除操作的实际情况，还提供了批量删除这样的接口，为之后的实验提供了便利。

而且这里我还碰到了 `js` 异步的问题，例如删除一条数据，给后端发请求删除指定数据并更新页面，后端可能还没来得及删就执行更新页面的操作了，导致更新页面后删除的那条数据仍然在页面中。解决方法就是将更新页面的操作放在回调函数中，删除操作返回后再更新页面。

三、前端

前端我采用 `vue` 框架，使用 `element-ui` 组件库、`axios` 网络请求库以及 `echarts` 可视化插件。前端的代码结构最重要的是那个 `components`，组件都在里面。因为不像有的同学那样套用 `GitHub` 上的优秀模板甚至沿用双创成果、本人也并非美工专业出身，一个个地去调试组件是疲惫且有些枯燥的，我在此就列举几个值得称道的事情来细讲一下吧。

为了实现全局的右键烟花特效，我在 `App.vue` 中加入了鼠标右键的监听事件。一有鼠标右键点击，就会有以点击点为中心，四散开来、分布均匀且颜色各异的圆形小球弹开，并且弹到一定距离之后就消失了。这个效果的实现主要源于在 `canvas` 画布上一步步的模拟，花了我不少的时间。

前端应用通过 `axios` 来监听 3000 端口（就是后端应用服务的端口）`request` 拦截器可以自请求发送前对请求做一些处理，比如统一加 `token`，对请求参数统一加密；`response` 拦截器可以在接口响应后统一处理结果，还能根据返回的是文件还是字符串灵活调整输出内容，比如说如果是文件就直接返回就好了，但如果是字符串数据就要先用 `JSON.parse` 进行打包

我选择 `Vue` 框架还有一个原因是因为它对响应式数据的支持非常好，这非常有利于进行动态的操作。数据模型仅仅是普通的 `JavaScript` 对象，当我们修改它们时，视图会进行更新，这使得状态管理非常简单直接。我在本次实验中实现的动态爬虫功能，只要对应的页面上点击一下“爬取最新新闻”按钮，就会向后端发送请求，然后开启爬取并存储进入数据库的流程，这就是一个响应式的数据之应用案例。

写 `Vue` 组件其实遵循的范式是类似的——`<template>` 标签内写文档大致的结构，就像是写 `HTML` 那样；`<script>` 标签内写组建的功能，我感觉主要是写 `methods` 内部的方法，就像写 `JS` 那样；最后再在 `<style>` 内调节样式，就像写 `CSS` 那样。总而言之，完成一个 `Vue` 组件的体验和开发原生的 `Web` 应用感觉类似，所以上手还是非常容易的。

作品效果图（整体、关键点和特效）

要求截取屏幕反映作品整体、关键点和特效的画面若干张，并插入本表中。

载入页面(背景是视频，动态播放哟)



最新新闻（新浪，网易，东方财富）

News 新闻管理系统

最新新闻

新浪国内新闻

网易国内新闻

东方财富

热搜看板

个人信息

关于

意见反馈

中国

请输入简介

请输入关键词

搜索

重置

新增

批量删除

爬取最新新闻

<input type="checkbox"/>	标题	简介	关键词	来源	操作
<input type="checkbox"/>	鲍十强正式担任中国男篮主帅，首秀11月亚运会主场迎战关岛	北京时间7月21日，中国篮协协会发布公告，为备战2027年男篮世界杯和2028年洛杉矶奥运会，结合未来国家男篮发展思路，根据中国篮协协会关于国家男篮主教练选聘方案...	中国男篮, 郭士强, 中国篮协, 国家男篮, 中国篮协协会	澎湃新闻	编辑 删除
<input type="checkbox"/>	中国气象局启动台风四级应急响应	来源：央视网央视网消息：据中国气象局网站消息，经综合研判和应急会商，中国气象局领导于21日9时30分签发命令，中国气象局办公室、应急减灾与公共服务司...	台风	央视	编辑 删除
<input type="checkbox"/>	国际奥委会：中国体育代表团征战十届夏季奥运会共获263金	中国体育代表团在夏季奥运会上总共夺得多少枚金牌？金牌数量最多的前十个项目是什么？在巴黎奥运会开幕前，国际奥委会给出了权威答案。	国际奥委会, 中国体育代表团, 夏季奥运会, 金牌, 奥运会	澎湃新闻	编辑 删除
<input type="checkbox"/>	中国篮协协会聘请郭士强担任中国男篮主教练	来源：中国篮协协会 为备战2027年男篮世界杯和2028年洛杉矶奥运会，结合未来国家男篮发展思路，根据中国篮协协会关于国家男篮主教练选聘方案，经研究...		新浪新闻综合	编辑 删除
<input type="checkbox"/>	文化中国行！“家门口”看古埃及文明 超95%文物亚洲首展	正在上海博物馆进行的古埃及文明大展，精选492组788件古埃及文明不同时期的珍贵文物，包括法老雕像、木乃伊棺，以及制作华美的金饰等...		央视新闻客户端	编辑 删除

共 79 条 5条/页 < 1 2 3 4 5 6 ... 16 > 前往 1 页

个人信息

News 新闻管理系统

最新新闻

新浪国内新闻

网易国内新闻

东方财富

热度看板

个人信息

关于

意见反馈

首页 个人信息

保存成功

用户名

test111

昵称

guo

邮箱

10211900416@stu.ecnu.edu.cn

电话

12345674567

地址

上海市普陀区中山北路3663号华东师范大学

新密码

确定

意见反馈面板

News 新闻管理系统

最新新闻

新浪国内新闻

网易国内新闻

东方财富

热度看板

个人信息

关于

意见反馈

首页 意见反馈

意见反馈

未来要结合所学继续添加功能呀~

提交

鼠标右键烟花特效（有渐变效果）



个人感想

伴随着期末大作业的结束，暑假短学期的 Web 编程课程画上了句号。因为项目开发的时间有限，我的期末项目还有很多进一步提升的地方，与自己最初的构想还是存在不小的差距。但这个项目麻雀虽小，五脏俱全，以前端为中心，依赖着后端、爬虫、数据库等部件，我对计算机科学领域“全栈工程师”的理念有更深刻的认知。在未来我还会学习分布式系统、云计算系统、人工智能这样的数据科学专业课程，凭借这些更深奥的知识，依托团队协作和项目管理，我相信今日 Web 编程奠定自己了日后开发超大型项目的基础。

炎炎夏日，老师比同学们更加辛苦。没有老师的悉心讲解，烦躁的大家没有精力理解繁琐的前端基本知识；没有老师的耐心指导，我们也不会完成如此的项目；没有老师的体贴陪伴，同学们在期末周之后无法静下心来学习三周。在此，我想向老师及助教表达最高的感谢！真的辛苦啦！