**江苏科技大学**

**课 设 报 告**

课 程： 信息内容安全与系统安全测试

学 院： 计算机学院

专 业： 信息安全

学 号： 182210710119

姓 名： 陈四贵

指导老师： 刘噶琼

设计时间： 2021-11-19至2021-11-20

# 课设目标

通过授课、实习等教学环节，使学生具有良好的工程职业道德、较强的社会责任感和较好的人文科学素养，并具有：

1、 信息获取和职业发展学习能力

2、 了解本专业从业人员的职能、信息内容安全与系统安全测试的专业发展方向；了解本专业的知识体系，以进一步明确该方向的目标。

3、 通过企业参观，了解企业最新的研究开发成果，并通过见习的手段将新成果应用到具体环境里，从感官到认知上明确信息安全类产品的需求及发展；

4、 通过企业小项目开发实践，了解具体项目的工作过程，掌握相关软件的安装及平台的搭建，通过讲授和实践初步了解项目的开发过程，部分同学能配合完成一个简单项目的开发工作。

5、 综合运用信息安全的理论、熟悉信息安全领域相关的技术标准、知识产权、产业政策和法律法规。

6、 具有较强的创新意识和进行产品开发和设计、技术改造与创新的初步能力。

7、 良好的质量、环境、职业健康、安全和服务意识，较好的组织管理能力、较强的交流沟通、环境适应和团队合作的能力。

# 课设内容

1、实验项目1·专家讲座

实验内容：邀请行业/企业专家，介绍本专业相关技术、职业发展方向、行业发展结构、技术标准等；介绍工程项目的开发流程、开发工具，以及相关的行业政策、法律和法规及行业工程师的业务素质要求及职业道德规范要求。

基本要求：了解本专业从业人员的职能、专业发展方向；掌握信息安全知识体系，以明确进一步本专业学习的目标和方向。

2、实验项目2·企业参观实习

实验内容：通过参观信息安全行业内企业，了解本专业的行业现状、产品和管理规范，熟悉信息安全方向最新的研究开发成果，并通过见习的手段将新成果应用到具体环境里，从感官到认知上明确行业产品的需求及发展。

基本要求：了解企业环境的基本概念；了解企业最新的研究开发成果，并通过见习的手段将新成果应用到具体环境里，从感官到认知上明确行业产品的需求及发展，了解具体项目的工作过程，了解信息内容安全与系统安全测试对企业的影响，明确行业给专业带来的生机。

3、实验项目3·职业道德教育、行业人才行为准则学习

实验内容：通过企业实习，了解企业环境的基本概念；体验企业文化，感受团队协作的责任和意义，以及让学生了解企业新开发产品的工作过程和现实意义。

基本要求：掌握行业需要的职业知识、基本技能，同时教育学生应具有良好的职业健康、职业安全和职业服务意识；能明确安全工程师的社会责任，理解网络空间安全对社会、知识和经济的影响和推动作用。

4、实验项目4·项目训练

实验内容：通过企业实习，完成企业导师提供一个简单工程实习项目，如网络安全管理、计算机病毒与防护、网络入侵检测、信息对抗与信息隐藏等，熟悉具体项目的工作过程，掌握相关软件的安装及平台的搭建，能够配合完成一个简单项目的开发工作。

基本要求：掌握各种信息安全技术，掌握相关软件的安装及平台的搭建，通过讲授和实践初步了解项目的开发过程，部分同学能配合完成一个简单项目的开发工作。

5、实验项目5·总结汇报交流

实验内容：能够根据查阅的资料和完成的实习项目，完成本次实习的报告，并能够汇报和交流。

基本要求：分组讨论、完成实习报告。

# 课设记录·专家讲座

出于课程设计需要，我们邀请了几位学院负责老师为我们介绍本专业相关技术及开发工具。通过这次课设我了解了很多。

首先，信息安全不是特定领域的安全，它无处不在，可以渗透入各个领域。

可以是军情领域。上世纪七十年代，信息技术刚刚兴起时，安全问题就已经凸显——通信时如何保证报文不被截获，亦或如何设计密码以最大程度遏止破译问题，这些情报（安全）问题在某种程度上就已经能决定战争的走向，在这个领域，信息安全技术的表现是信息隐藏技术。

可以是民生领域。在高度现代化的这个时代，很多家庭都在家里安装了摄像头等工具，追根溯源，这些工具也是要通过计算机网络起作用的。一些不法分子就会通过某些非法手段，入侵摄像头，获取人的隐私以攫取非法利益。在这个领域，信息安全技术的表现是物联网安全以及计算机网络安全。

可以是军事领域。如今各国军事都已高度发达，依靠卫星的大规模杀伤性武器占了很大一部分。在军事中，能通过各种技术来干扰敌方设备，以达到使对方武器失效，进而赢得战争的目的。在这个领域，信息安全技术的表现是通信安全以及信息对抗技术。

其次，在实践中，我了解到了网络攻防工具方面的问题。我了解到，很多工具其实都是利用协议本身的特征进行信息判断、漏洞利用的。针对这些，在已有工具Httprint的基础上，我从原理上构思了一个全新的工具，如下所示。

其名称为基于机器学习的智能渗透测试工具SP-work。

SP-work是基于现有平台测试工具，在广泛测试和实践的基础上，汲取经验所开发的自动化、智能化渗透扫描工具，具有广泛的实际意义。通过伪装，它能做到智能爬取目标站点数据， 结合Httprint的相关原理，构造签名引擎库，获取目标服务器信息。

在设计上，SP-work采用著名的MVC设计模式，将前端界面、后端逻辑分隔开来，结构清晰，有利于开发。

具体来说，由于Java跨平台、可移植、面向对象、功能强大、易于使用的特点，加之许多渗透工具由Java编写，直接或间接地需要Java环境运行，主要考虑到可移植性和软件复用，我们采用Java语言进行开发。

Java FX技术主要应用于创建Rich Internet Applications，功能也较为强大，较之Java Swing也毫不逊色。由于我们使用了MVC设计模式，因而，我们将会依托Java FX中的M-model、V-fxml、C-controller进行设计。

由于Java FX有专用的UI设计工具Scene Builder，我们采用其进行前端界面设计。与大部分软件类似，SP-work拥有类似的界面——主界面铺叙，功能菜单阵列在旁。各项功能依托功能菜单开展、实现。

设计中，基面板采用SplitPane，设有ScrollPane、MenuBar等。

类似爬虫，SP-work会利用ICMP与TCP/IP协议，向目标服务器发送请求，获取信息。若不能正常爬取信息，则通过设定规则与返回的错误，控制浏览器返回数据包，比对差异，利用机器学习，智能化设置自身参数，以达到伪装浏览器，欺骗目标服务器、爬取信息的目的。其中，主要是对自身设置代理、数据报头的调整以及对所爬取数据包的解析，若能正常爬取信息，则进入下一流程。

下一流程是对所获取数据的分析，这一流程与Httprint的工作流程类似，都是分析服务器信息。基于上述这些返回信息的差异，Http会为服务器生成十六进制编码的ASCII码串，即服务器返回信息的一种表现形式。SP-work先把一些HTTP签名信息保存在一个文档里，然后分析那些由HTTP服务器产生的结果。当发现那些没有列在数据库中的签名信息时，可以利用SP-work产生的报告来扩展这个指纹签名数据库，在下次运行SP-work时，这些新加的签名信息也就可以使用了。有了签名信息的文档之后，再运用统计学原理和模糊逻辑技术，SP-work能很有效地确定HTTP服务器的类型。识别算法工作原理：第一步，把所有已知类型的服务器的签名信息全部加载为签名集S（比对数据库）；第二步，然后对测试服务器运行指纹引擎，得到返回签名SR,换句话说，也就是对一个未知服务器进行指纹测试后，返回这个服务器的签名信息SR；第三步，这里是将未知服务器的签名信息SR和签名集中的签名信息逐个比对，得到权值wi（未知服务器和已知服务器签名信息的相似度）；第四步，通过模糊逻辑函数，一一求出签名si在整个签名集中最匹配的概率，即置信度。

在确认服务器信息之后，SP-work并不会停止工作，它会在用户选择的前提下，利用该版服务器对应的漏洞，对目标服务器进行渗透。也就是说，它能获取服务器信息，并根据这个信息找到目标服务器的漏洞，进行相应的渗透，以获取更多的信息。

在所有分析-渗透-分析阶段结束之后，SP-work会综合获取到的所有的信息，将其导出为报告供用户查看，主要包括通信双方使用的协议、通信端口、数据流图、服务器版本、服务器对应的漏洞等信息。

对HTTPrint而言，若不能正常收到回传的报文，则无法正常扫描。相较于Httprint，SP-work能充分利用爬虫的优势，伪装成浏览器，获取信息，这保证了扫描工作的正常进行，有着划时代的重要意义。此外，SP-work中存放有各版本服务器的漏洞信息，较之以往的Httprint更为全面。

随着信息化的不断推进，安全形势已成为一个越来越不可忽视的问题。这需要我们去投入、解决它。

# 课设记录·企业参观学习

为进一步增强自身专业技能，促进达到现代企业的能力要求，进一步适应现代工业标准，出于工作需要，我针对校学工系统研究了爬虫的使用，同时对现代企事业单位的信息化水平有了一定的了解。

写爬虫，需要做的就是分析浏览器的行为，并模拟浏览器发送请求。

下面我们通过详细分析浏览器的行为，实战模拟爬取江苏科技大学信息门户及其子网站。

打开谷歌浏览器，进入“江苏科技大学信息门户”。

输入错误的账号或密码，点击登录，分析浏览器行为。由于登录请求命名中一般含有Login、register等字样，且对于网站源码的请求一般在Network中的DOC类型（其他类型返回的不是html页面）。我们重点关注DOC类型的日志。如图4-1所示。

分析后，发现DOC类型有一个含有login的post请求，其通过form data提交了我们输入的账号、密码，可以确认，这个就是我们要模拟的请求。如图4-2、4-3所示。

但是，该post请求中提交的表单数据里不仅仅有我们已知的username以及password，还有execution、\_eventId、loginType、submit等字段。初步考虑，这可能是通过get该登录页面自动获取的东西。

查看网页源码，发现这些字段都属于网页的属性，是固定的，不是动态获取的。如图4-4所示。

由于post请求需要提交这些信息，我们在提交post请求前需要获取这些字段。这就有两个方向，一是查看网页源码，将这些字段直接复制，封装到header里；另一个方向是通过get请求获取网页源码，再经过Beautifulsoup等库进行数据处理动态获取这些信息。这里我们采用后者。

|  |
| --- |
| response = self.\_\_session.get(url=url\_menhu, headers=headers1, verify=False)  dic = {}  # 获取表单中的一些参数  lt = BeautifulSoup(response.text, 'html.parser')  dic['execution'] = lt.select('input[name="execution"]')[0].get("value")  dic['\_eventId'] = lt.select('input[name="\_eventId"]')[0].get("value")  dic['loginType'] = lt.select('input[name="loginType"]')[0].get("value")  dic['submit'] = lt.select('input[name="execution"]')[0].get("submit")  dataMenhu = {  'username': self.\_\_username,  'password': self.\_\_password,  'execution': dic['execution'],  '\_eventId': dic['\_eventId'],  'loginType': dic['loginType'],  'submit': dic['submit']} |

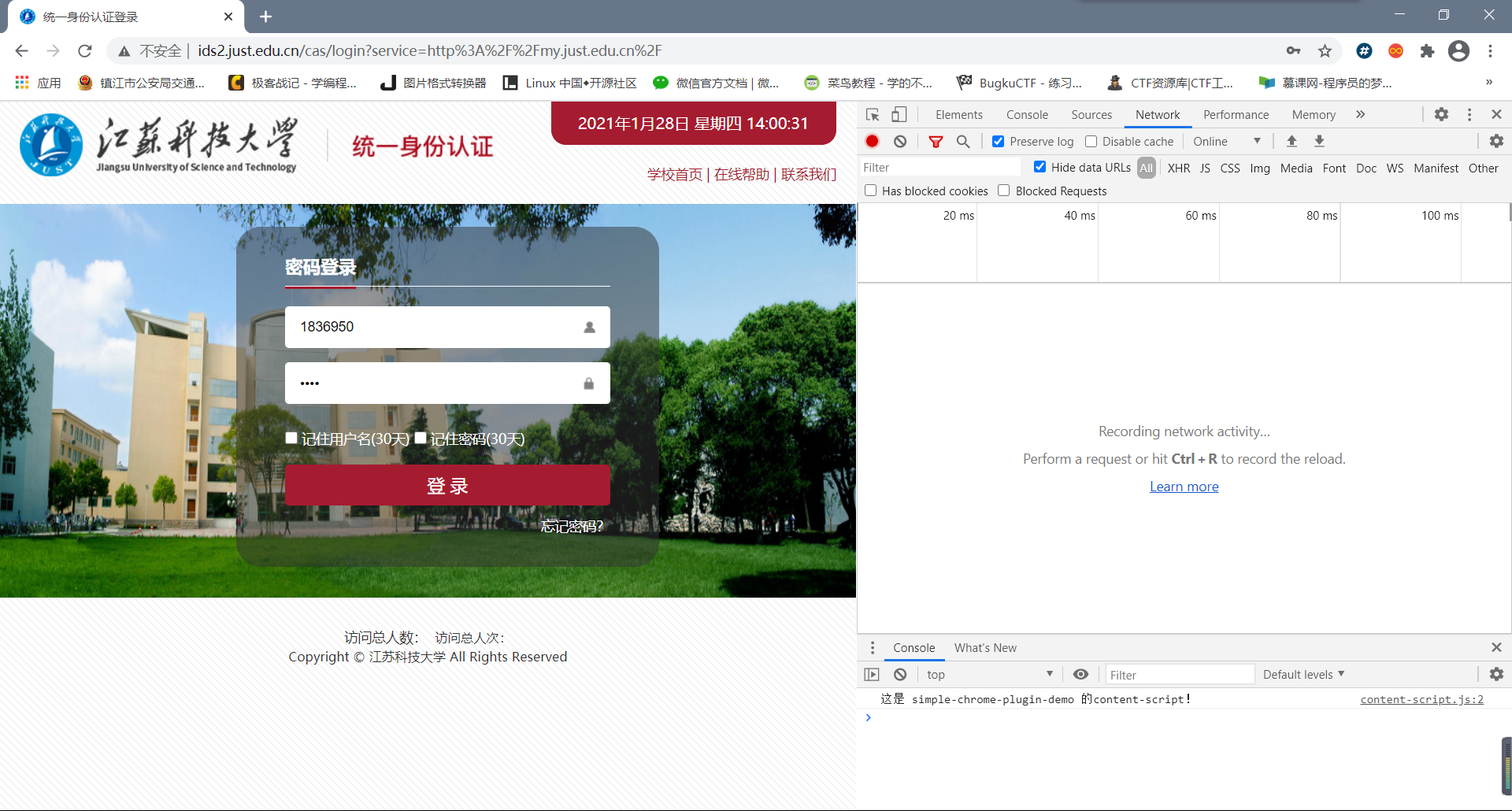
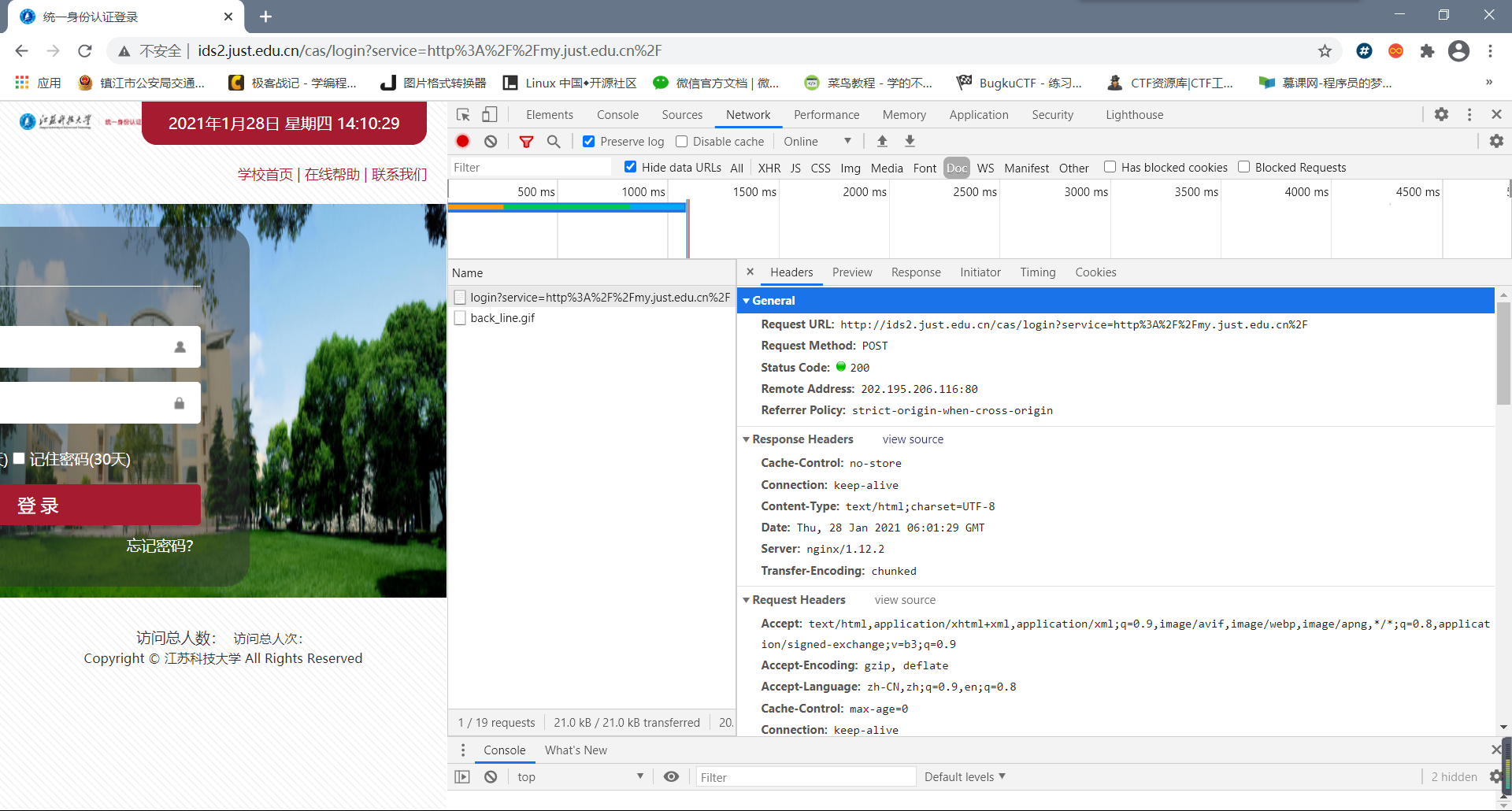
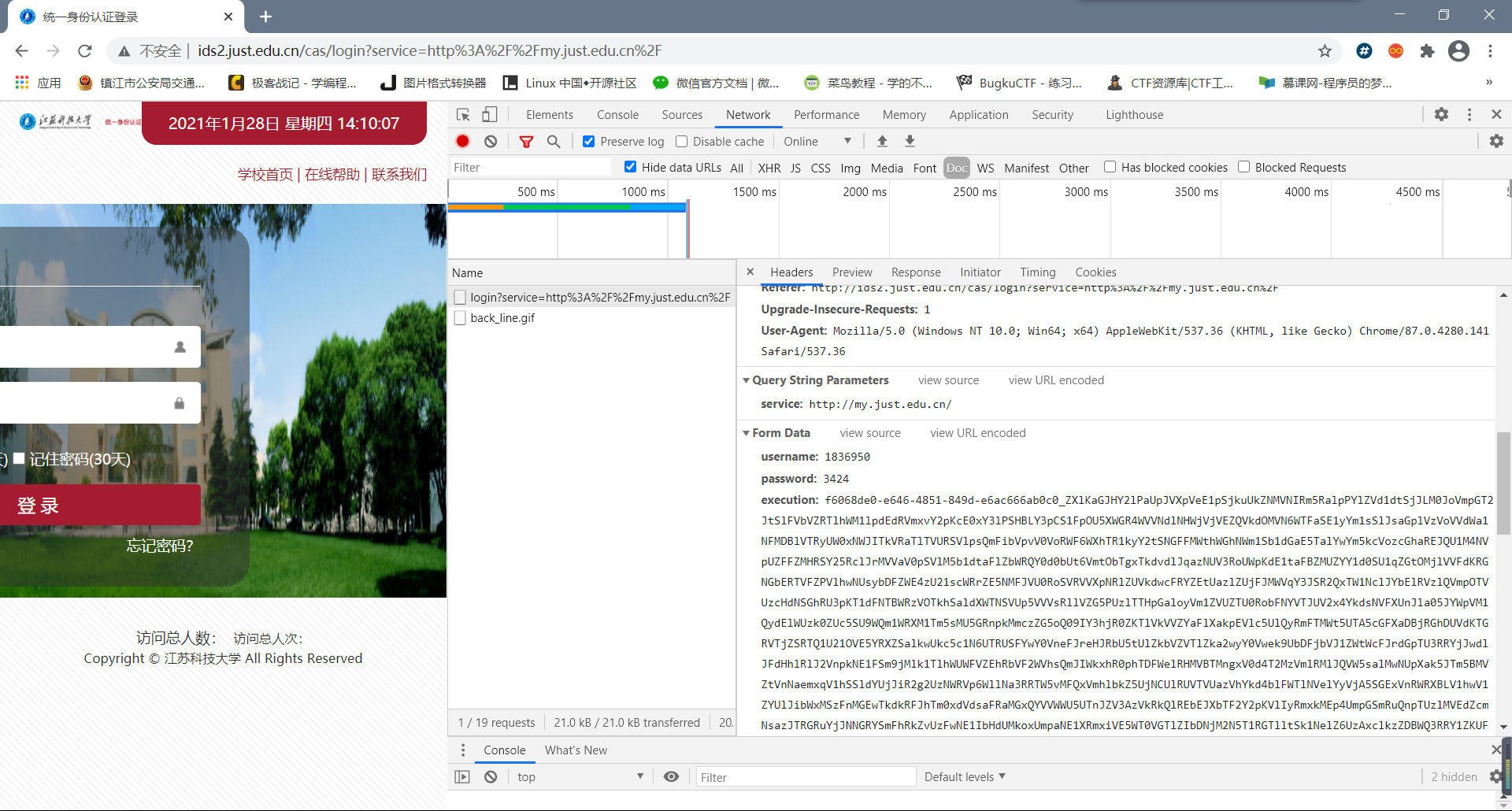


图4-1

图4-2



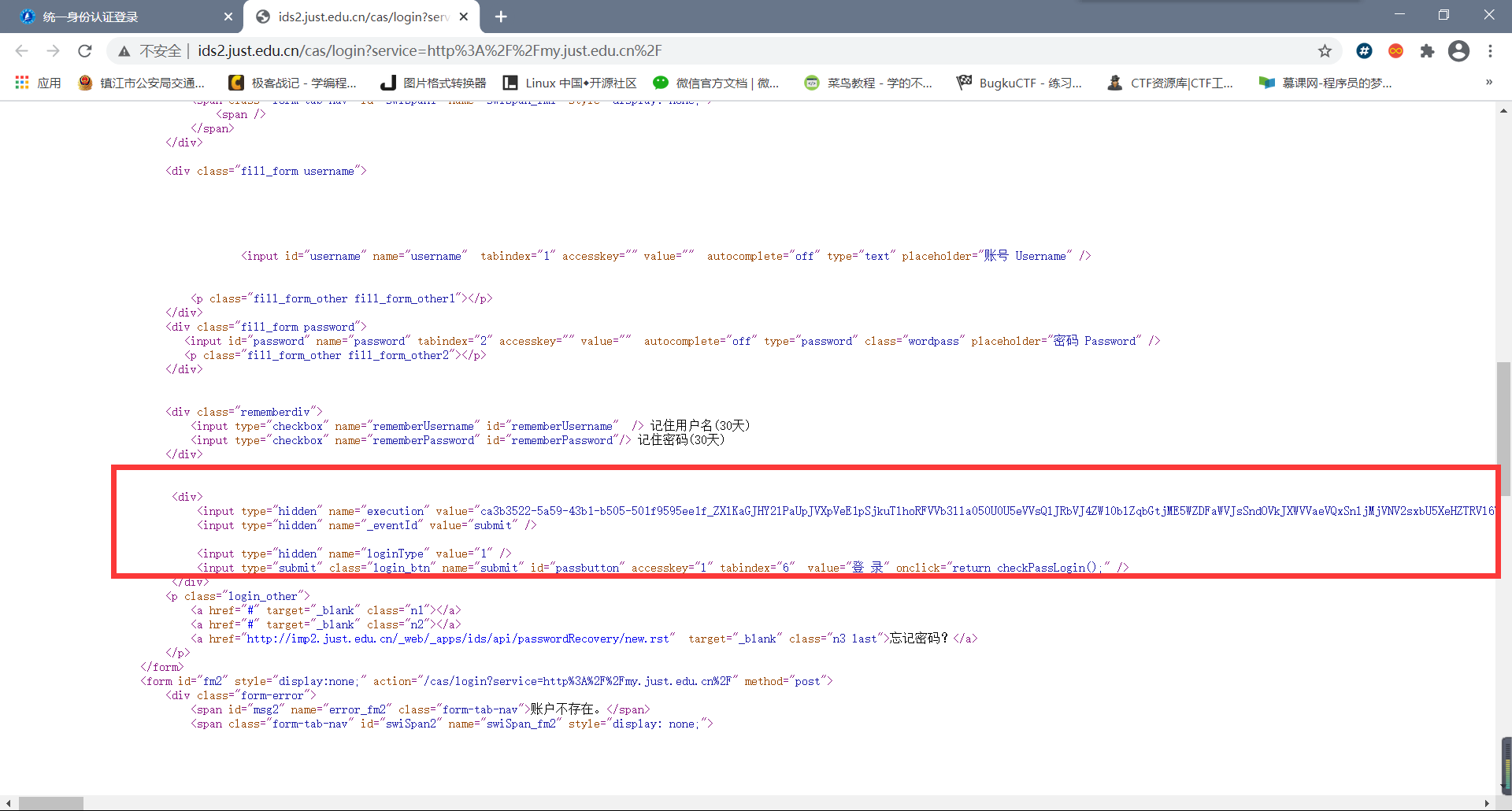
图4-3

图4-4

获取应该post的参数后，模拟浏览器发送post登录请求

|  |
| --- |
| headers1 = {  'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/87.0.4280.141 Safari/537.36'}  response = self.\_\_session.post(url=url\_1, data=dataMenhu, headers=headers1, verify=False) |

我们可以通过响应的内容验证是否成功登录，可以看到，若成功登录，返回页面不会显示“统一身份认证登录”。由此作为成功登录的标志。

|  |
| --- |
| if "统一身份认证登录" in response.text:  print("{}登录门户网站失败\n".format(self.\_\_username))  return False, self.\_\_session  else:  print("{}登录门户网站成功\n".format(self.\_\_username))  return True, self.\_\_session |

成功登录后，分析请求，可以发现若干与login相关的DOC请求。点开查看后，发现每个响应头里都有set-cookie字段。如图4-5所示。

图4-5

可以判断，该网站使用的是动态cookie，登录成功后服务器会分配给客户端新的用于验证的cookie。如果不更新cookie而直接访问子网站，会返回登录页面的源码。因此，虽然这些名字带有Login的申请页面并不是我们最终所需要的，我们也需要使用同一个session发送get请求，由此更新我们的cookie，用于访问子网站。

|  |
| --- |
| session.get(url=url\_2, headers=headers1)  session.get(url=url\_3, headers=headers1)  session.get(url=url\_4, headers=headers1)  session.get(url=url\_5, headers=headers1)  session.get(url=url\_6, headers=headers1)  session.get(url=url\_7, headers=headers1)  session.get(url=url\_8, headers=headers1) |

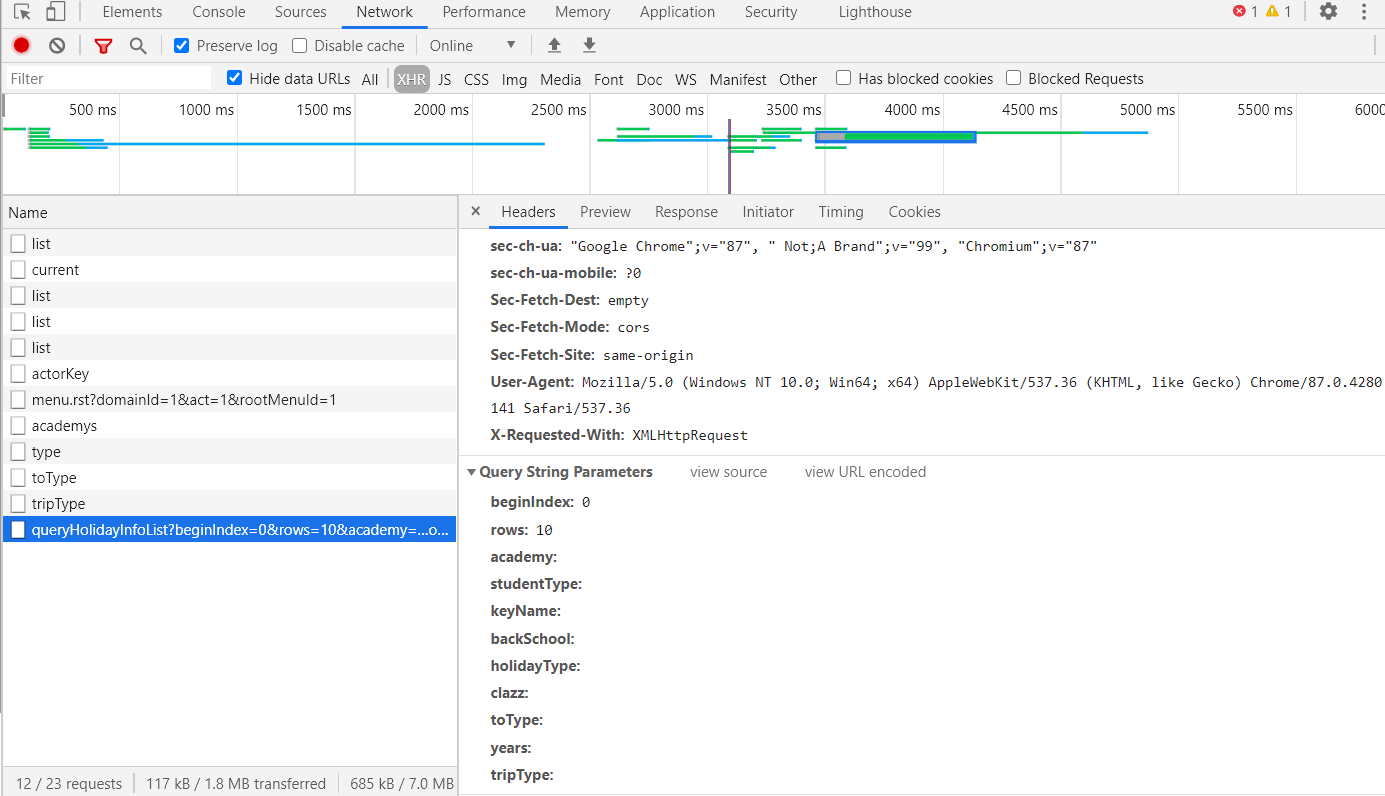
动态更新cookie后，通过get发送请求至子系统网址，成功登录子系统。可以看到，随着页面的滑动，子系统的网址始终没有发生变化，而显示的数据有所不同（rows=10时显示10条数据，rows=20时显示20条数据），这表明该网站使用了ajax技术，我们需关注XHR日志，以此行为为据按照我们的要求编写程序，获取数据。如图4-6所示。

图4-6

此外，根据认识论的观点，认识的过程通常是认识-实践-再认识，这次实践也不例外。由于工作需要，我需要每月一次运行该程序，但在此后的工作中，几乎我每次运行该程序，都会出现新问题——也就是无法正常爬取数据。这体现了信息技术的发展和企事业单位安全性的提升。在寻找bug、改进程序的过程中，我意识到，一般的企事业单位不会设置太多反爬虫机制，通常情况下只是优化服务器配置，识别更多非浏览器行为的请求——很多时候爬取不出数据的原因仅仅只是请求头少了某个真实的字段，和真实浏览器请求有偏差导致的。

这次实践一方面加强了我的专业技能，另一方面，也使我对现代企事业单位的信息化水平有了一定的认识，再一方面，也使我认识到了信息内容安全和系统安全测试的重要性。同时，我也会根据自身经验，在未来的学习生活工作中致力于安全建设。

# 五、课设记录·职业道德教育、行业人才行为准则学习

当前网络安全问题不容忽视。但在日常生活中，有许许多多的企事业单位并没有认识到安全的重要性，也没有意识到自己的行为是违法行为，从而引咎入狱，实在是很不应该。

2017年5月9日下午，最高人民法院、最高人民检察院《关于办理侵犯公民个人信息刑事案件适用法律若干问题的解释》正式颁发。这确是一部具有里程碑式意义的法律，也是互联网公司关于个人信息、网络安全及数据的合规大法。其体现在以下几个方面。

其一，个人信息的扩大化界定。

作为立法和司法领域之中关于个人信息的顶级位阶，《网络安全法》和两高新解对个人信息的界定不尽相同。相比较而言，两高新解关于个人信息的认定可具扩展性，故保护更近严格。这种扩展性主要体现在如下维度。

两高新解增加了活动情况这一内涵，即类如自然人的行踪轨迹也界定为一种个人信息，而不仅仅只是个人隐私。

其二，网络浏览记录属于个人信息吗？

除了GPS技术外，网络cookie或网络信标臭虫技术，其实也是能给用户画像，并反映出个人的行踪轨迹的，它能把个人网页浏览记录保存下来，并和个人网络账号例如某特定浏览器等一一结合，从而识别特定虚拟人的特定活动。这类情况下的行踪轨迹似科更接近个人信息的定义。

中国隐私权纠纷第一案中，一二审法院作出了截然相反的判决。南京中院二审的终审判决，对利用cookie技术收集的用户上网活动轨迹信息是不是个人信息作出了明确的回答——上网轨迹不属于个人信息。

理由如下：

1.百度网讯公司在提供个性化推荐服务中运用网络技术收集、利用的是未能与网络用户个人身份对应识别的数据信息，该数据信息的匿名化特征不符合“个人信息”的可识别性要求。

2.网络用户通过使用搜索引擎形成的检索关键词记录，虽然反映了网络用户的网络活动轨迹及上网偏好，具有隐私属性，但这种网络活动轨迹及上网偏好一旦与网络用户身份相分离，便无法确定具体的信息归属主体，不再属于个人信息范畴。

3.百度网讯公司个性化推荐服务收集和推送信息的终端是浏览器，没有定向识别使用该浏览器的网络用户身份。虽然朱烨因长期固定使用同一浏览器，感觉自己的网络活动轨迹和上网偏好被百度网讯公司收集利用，但事实上百度网讯公司在提供个性化推荐服务中没有且无必要将搜索关键词记录和朱烨的个人身份信息联系起来。

这里就涉及到网络虚拟人的问题，法院判决书中第3点理由的意思指，运用cookie技术在推荐广告时，它推送的对象其实不是某个账号所对应的真实自然人，而仅仅是某个软件或设备，例如当A使用其在家里的台式电脑，并使用该电脑上的火狐浏览器浏览袜子时，cookie技术下次就会向这台电脑的这个浏览器即虚拟人再行推送类似商品，而不会向A这个人的其它设备推送。因为cookie技术仍是一种本地化技术实现。

但若cookie技术通过祁同伟登陆的火狐浏览器帐户同一性认定，跨终端平台推送广告，似乎就需要和个人身份账号密码相联系了，这点又有不同，这种功能在两高新解下就可能变成灰色地带了。

2016年6月底，谷歌把2007年曾向用户承诺的隐私条款不动声色地拿掉了：“除非用户同意，谷歌不会将下属广告公司DoubleClick收集的 cookie信息，和谷歌收集的其它个人身份信息结合起来使用”。说明这种cookie技术的效果还将继续，而且也会发挥越来越大的作用。

其三，个人隐私和个人信息的边界。

个人信息和个人隐私是有区别的，不能完全等同。个人信息包括部分个人隐私，但个人隐私不一定就是个人信息；个人隐私除了用户个人信息外还包含私人活动、私有领域例。个人隐私和个人信息的边界需要进一步界定。

根据王利明教授的观点，个人隐私和个人信息的区别还是非常大的。主要是三点，大意如下：第一，隐私权是一种精神性人格权，财产价值不突出。而个人信息性质上属于一种集人格利益与财产利益于一体的综合性权利，对于一些名人的个人信息而言，甚至主要体现为财产价值。第二，隐私权是一种消极的、防御性的权利，只能在遭受侵害的情况下请求他人排除妨害、赔偿损失等。而个人信息，权利人除了被动防御第三人的侵害之外，还可以对其进行积极利用，如出售。第三，隐私不限于信息的形态，它还可以以个人活动、个人私生活等方式体现，且并不需要记载下来。而个人信息必须以固定化的信息方式表现出来，因此，个人信息通常需要记载下来，或者以数字化的形式表现出来。也就是说，个人信息概念侧重于“识别”，即通过个人信息将个人“识别出来”。

以往，侵犯个人隐私一般很难上升到刑事犯罪，主要由侵权责任法进行私力求济。而侵犯个人信息却不同，不论是治安管理，还是现在的刑事犯罪都可以寻求帮助，在民法总则出来后，个人更是可以通过私力救济来保护个人信息了。这也足说明在保护法益上，个人信息的要显得更加重要一些。也正是因为此，我们更要注意区别两者，既不能把侵犯个人隐私的事，用刑法加以打击，也不能把侵犯个人信息的犯罪，不予受理而推诿扯皮。

总结来说，虽然《最高人民法院、最高人民检察院关于办理侵犯公民个人信息刑事案件适用法律若干问题的解释》颁布，但是，最为重要的问题，或者说最容易在打击犯罪中搞错的，不是5000条个人信息就入刑的问题，而是个人信息的界定问题。

在刑事辩护中，公安机关所认定的每一条所谓个人信息或数据，都值得掰开了揉碎了好好匹配研究，每一条信息都必须是能指向一个特定人才视为有效信息，这是定罪的前提，否则就变成纯粹的迎合民愤的眉毛胡子一把抓了。这也诚如两高新解中所述的：经过处理无法识别特定个人且不能复原的除外。

对于互联网企业来说，随着网络安全法和两高新解等出台，个人信息保护要求更高了，若网络服务提供者拒不履行法律、行政法规规定的信息网络安全管理义务，经监管部门责令采取改正措施而拒不改正，致使用户的公民个人信息泄露，造成严重后果的，应当依照刑法第二百八十六条之一的规定，以拒不履行信息网络安全管理义务罪定罪处罚。

因此，信息安全是十分重要的，目前国家也很重视这一领域。在日常学习的过程中，我们也要分清界限，以避免引起不必要争端。

# 六、课设记录·项目训练

随着信息技术的发展，信息化、数字化已成为社会发展不可逆转的趋势。这种趋势也给我们带来了很多的安全隐患。诸如爬虫等信息技术，就能获取到很多隐私信息。

现针对爬虫开展实践。

◎第一步，打开目标网站，利用Chrome自带的Network工具捕获数据包。如图6-1、6-2所示。

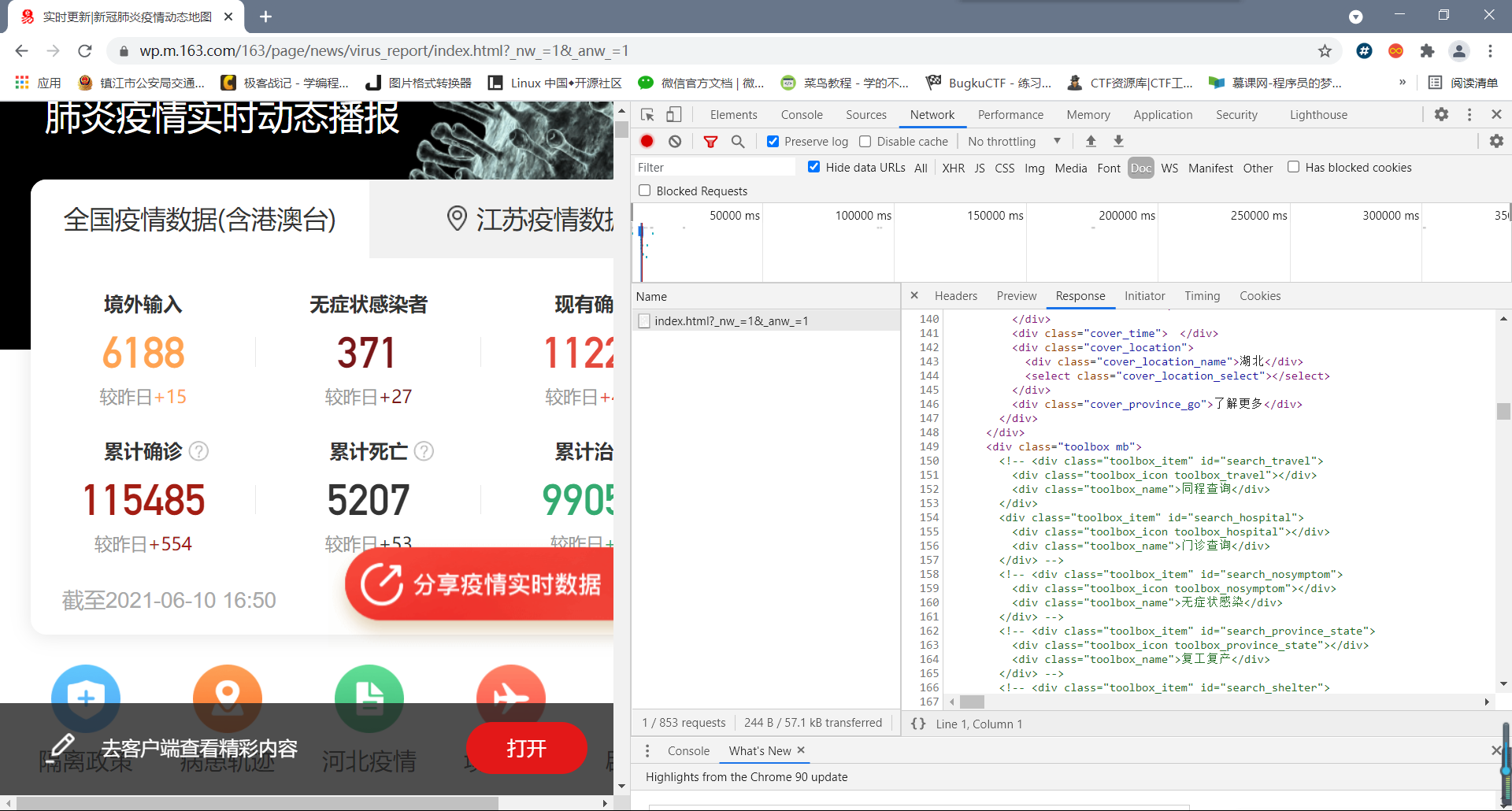
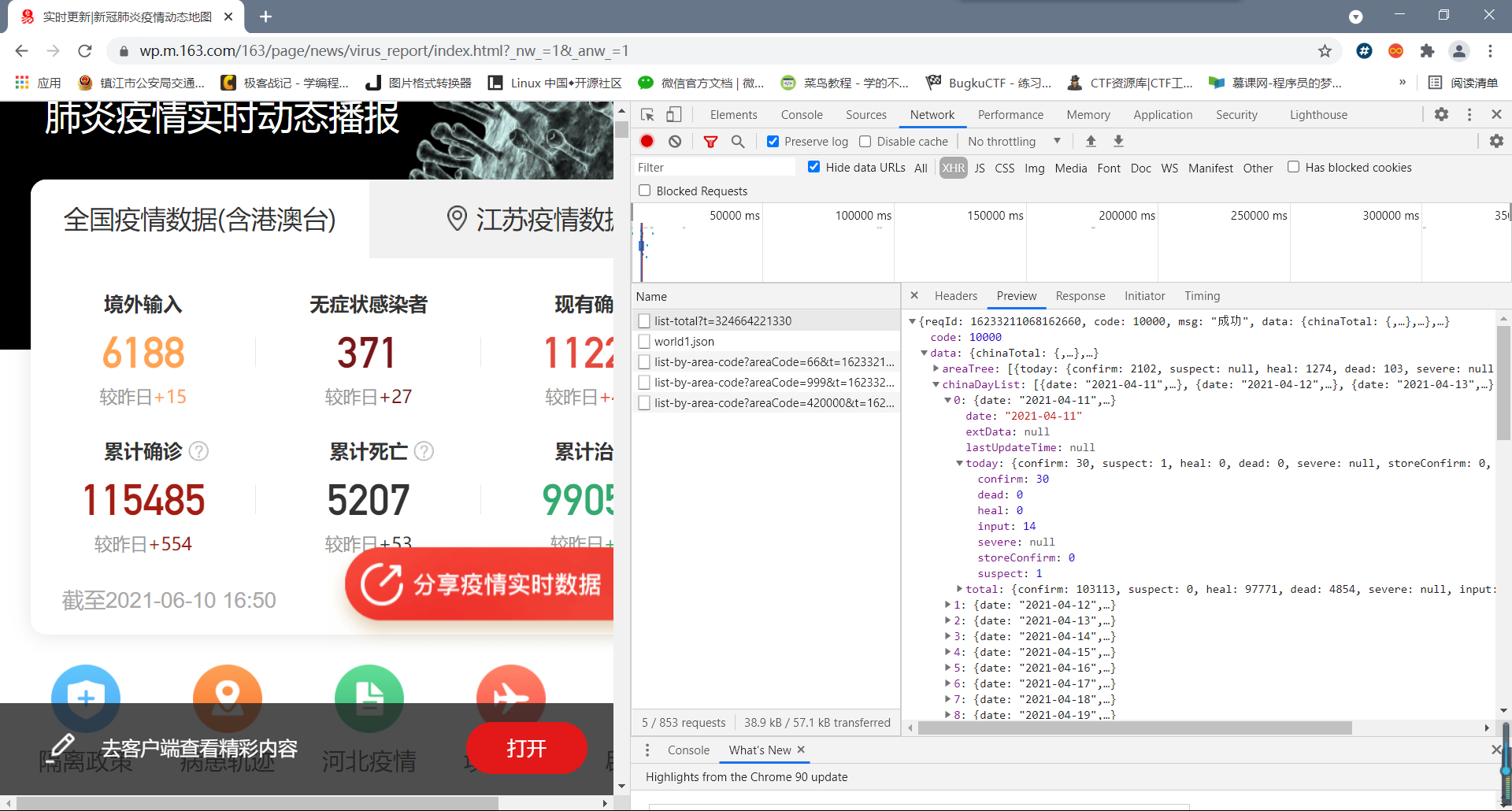


图6-1 捕获数据包

图6-2 捕获数据包

分析后发现，该网页使用了AJAX技术，是一个动态网页。它的数据与页面基本式样是分开的，html页面只是一个空壳，数据通过AJAX技术异步传输。因此，我们若要得到想要的数据部分，需要重点关注XHR请求部分。

◎第二步，确定所需信息，比对特征以锁定目标数据包。

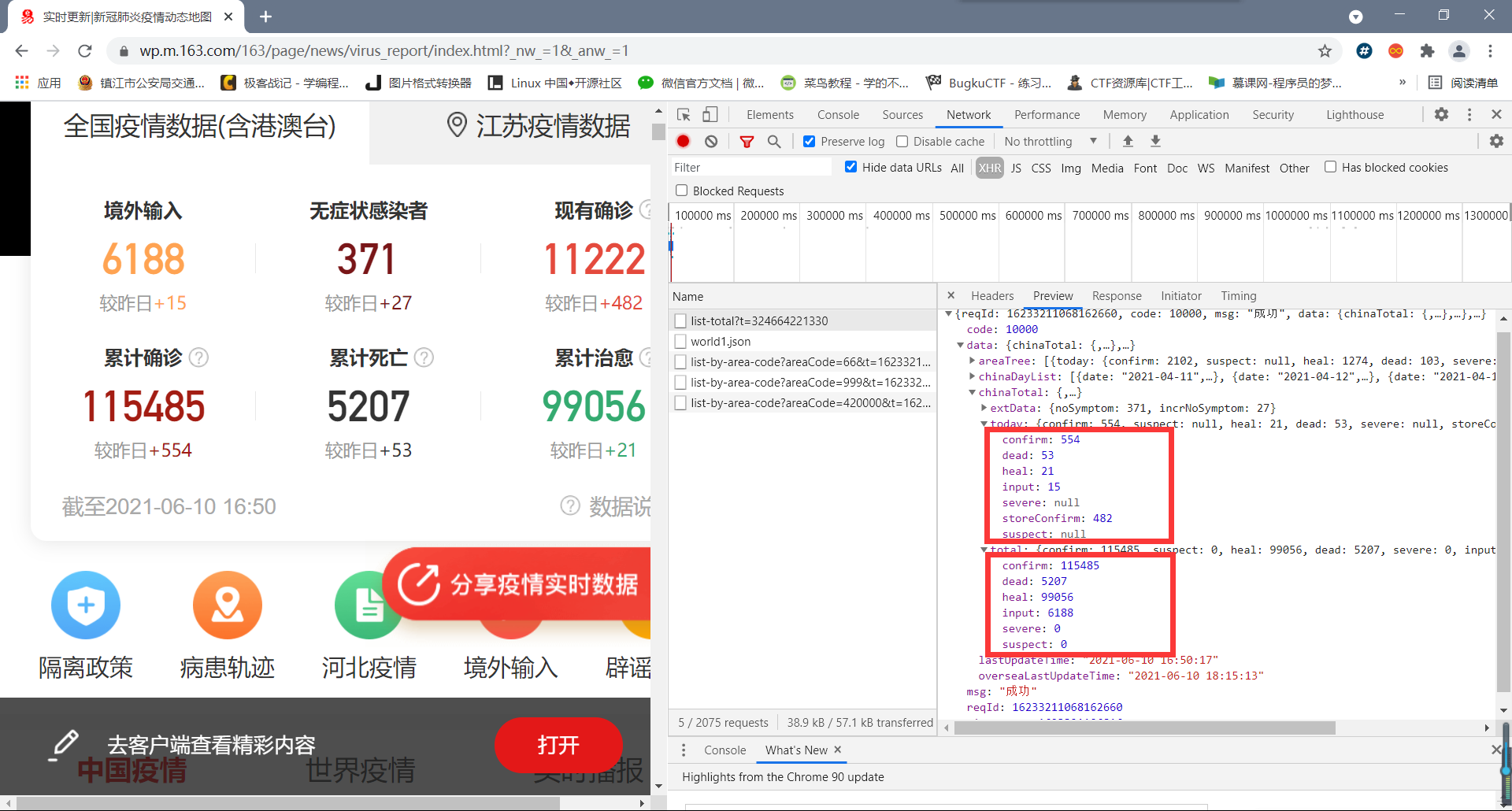
假定我们需要总体情况数据（即当前境外输入、无症状感染者、现有确诊等信息及其当日新增数），则根据数据比对，迅速确定“list-total”数据包为我们所需要的。如图6-3所示。

图6-3 目标数据包内容

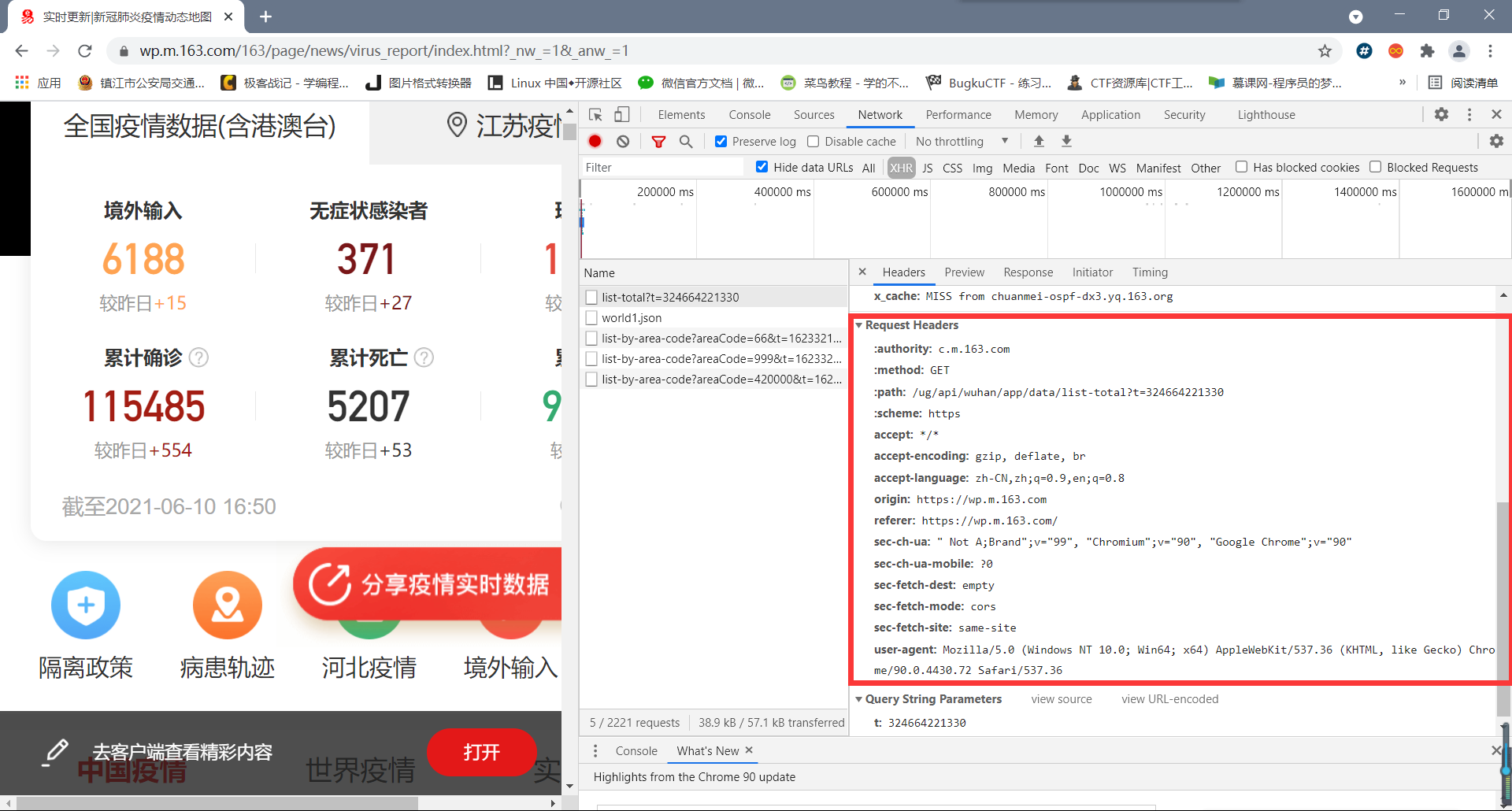
◎第三步，根据数据包报头内容，编写请求头以伪装浏览器，达到欺骗服务器的目的。如图6-4。

图6-4 目标数据包报头信息

由此编写请求头为：

1. headers1 = {
2. 'accept': '\*/\*',
3. 'accept-encoding': 'gzip, deflate, br',
4. 'accept-language': 'zh-CN,zh;q=0.9,en;q=0.8',
5. 'origin': 'https://wp.m.163.com',
6. 'referer': 'https://wp.m.163.com/',
7. 'sec-ch-ua': '" Not A;Brand";v="99", "Chromium";v="90", "Google Chrome";v="90"',
8. 'sec-ch-ua-mobile': '?0',
9. 'sec-fetch-dest': 'empty',
10. 'sec-fetch-mode': 'cors',
11. 'sec-fetch-site': 'same-site',
12. 'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/90.0.4430.72 Safari/537.36'
13. }

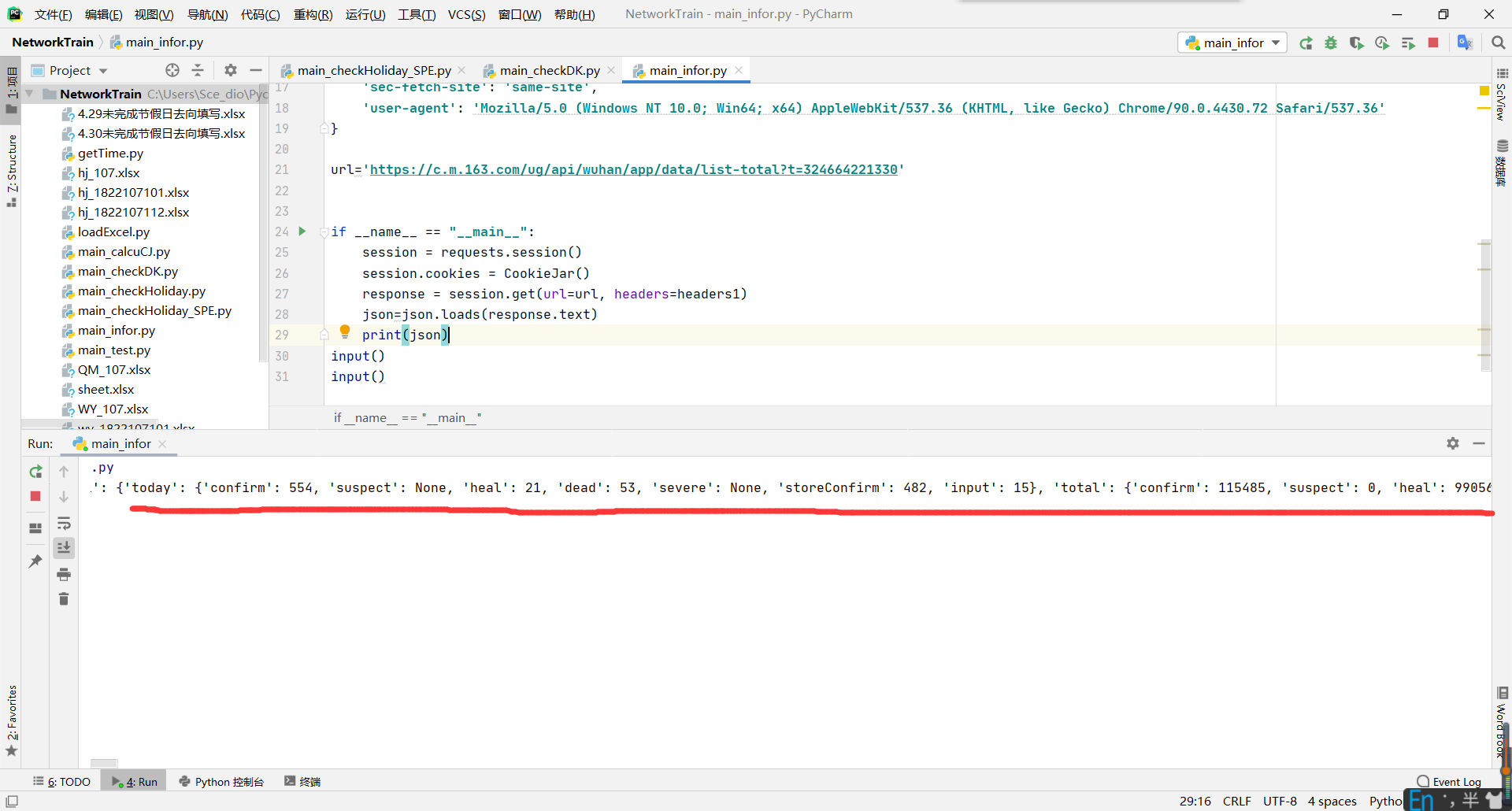
使用json加载后，比较获取的报文内容与我们需要的数据，发现与我们想要的相符。获取成功。如图6-5.

图6-5 获取到的报文内容打印

◎第四步，处理json数据，将其保存到excel表格中，以便后续分析。如图6-6.

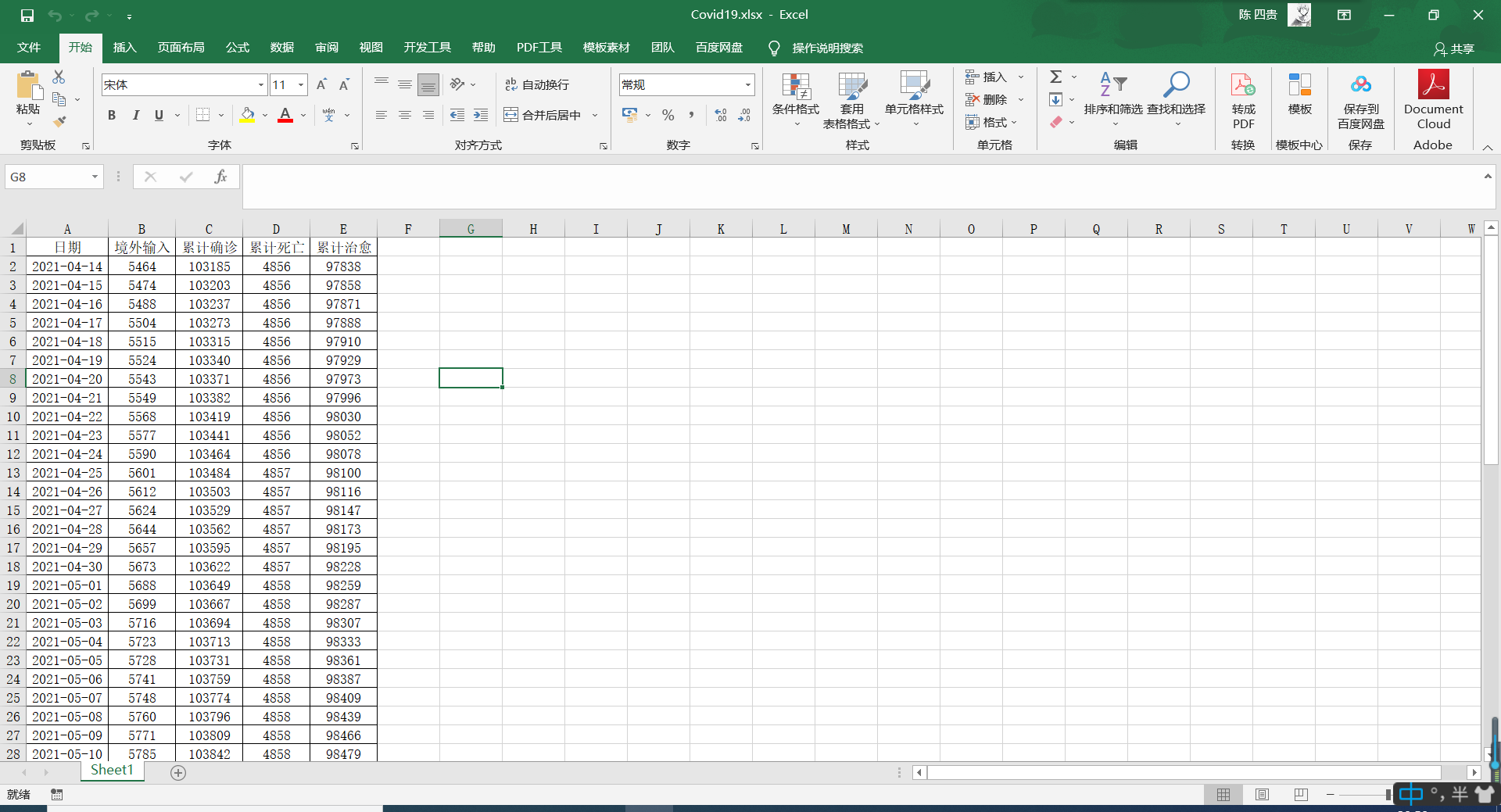
1. f = Workbook()
2. sheet = f.active
3. sheet.title = "Sheet1"
4. row0 = ["日期", "境外输入", "累计确诊", "累计死亡", "累计治愈"]
5. sheet["A1"].value = row0[0]
6. sheet["B1"].value = row0[1]
7. sheet["C1"].value = row0[2]
8. sheet["D1"].value = row0[3]
9. sheet["E1"].value = row0[4]
11. counter = 2
12. **for** i **in** range(len(json)-1):
13. json\_1=json[i]
14. sheet["A" + str(counter)].value = json\_1['date']                         #日期
15. sheet["B" + str(counter)].value = json\_1['total']['input']               #境外输入
16. sheet["C" + str(counter)].value = json\_1['total']['confirm']             #累计确诊
17. sheet["D" + str(counter)].value = json\_1['total']['dead']
18. sheet["E" + str(counter)].value = json\_1['total']['heal']
19. counter = counter + 1
20. f.save("Covid19.xlsx")

图6-6 数据预处理后的excel表格

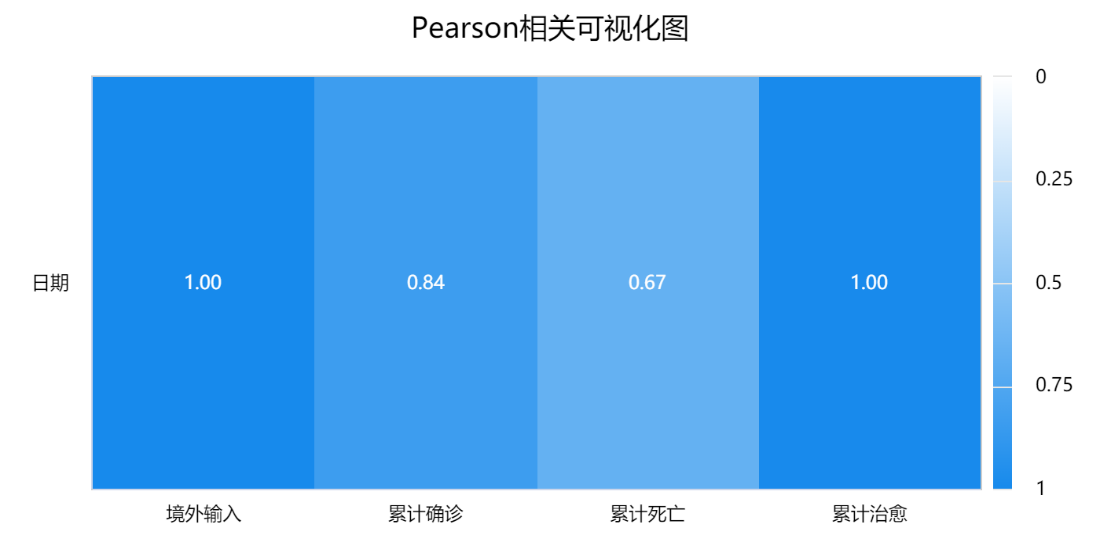
◎第五步，对相关数据进行数据挖掘，找出所需信息。

由于我们预处理后的信息是新冠疫情下各项数据随时间的变化情况，因而我们应主要研究数据的变化趋势，根据数据挖掘判断数据的变化趋势、相关性等。

首先可给出各项数据变化趋势折线图，如图6-7所示。

由图中数据可以看出，发生疫情以来，大部分病例是本土病例。在疫情防控常态化下的如今，国内疫情形势已基本得到控制，但仍存在部分新增本土病例和境外输入的问题。此外，随着我国医疗压力的逐步缓解，对新冠肺炎患者的救助能力也在上升，康复率有显著提高。种种数据告诉我们，即使疫苗已经普及，我们仍然不能放松警惕，仍要保持疫情防控常态化，坚决把疫情的势头扼杀在摇篮中，这样才能很好地控制疫情。

然后通过相关分析，我们可以用相关数据科学地描述变化趋势。



从上表可知，利用相关分析去研究境外输入、累计确诊、累计死亡、累计治愈分别和日期共1项之间的相关关系，使用Pearson相关系数去表示相关关系的强弱情况。具体分析可知：

境外输入与日期共1项之间全部均呈现出显著性，相关系数值分别是0.999，并且相关系数值均大于0，意味着境外输入与日期共1项之间有着正相关关系，这可能意味着国外的疫情愈演愈烈。累计确诊与日期共1项之间全部均呈现出显著性，相关系数值分别是0.837，并且相关系数值均大于0，意味着累计确诊与日期共1项之间有着正相关关系。累计死亡与日期共1项之间全部均呈现出显著性，相关系数值分别是0.668，并且相关系数值均大于0，意味着累计死亡与日期共1项之间有着正相关关系。累计治愈与日期共1项之间全部均呈现出显著性，相关系数值分别是0.997，并且相关系数值均大于0，意味着累计治愈与日期共1项之间有着正相关关系。

借由此次实践，我有以下几方面感受：

1、数据获取：爬虫是很有效的获取数据的一种方式，但若应用不当，也可能违反法律。在实际爬取的过程中，应注意结合实际分析数据包。比如，可充分利用Chrome Network工具，比对我们所需要的数据，找到数据包。此外，有的网页是静态网页，所有的信息都在html中体现，但有的网页使用了AJAX技术，其网页和数据异步加载。对于这样的网站，我们应着重关注XHR请求，因为我们想要的数据通过此种方式呈现。

2、数据挖掘：数据挖掘是指从大量的数据中通过算法搜索隐藏于其中信息的过程，是人工智能和数据库领域研究的热点问题。在这样的大数据时代，我们应当掌握数据挖掘算法。

设计使用源代码如下所示：

1. **import** requests
2. **import** json
3. **from** http.cookiejar **import** CookieJar
4. **from** openpyxl **import** Workbook
6. headers1 = {
7. 'accept': '\*/\*',
8. 'accept-encoding': 'gzip, deflate, br',
9. 'accept-language': 'zh-CN,zh;q=0.9,en;q=0.8',
10. 'origin': 'https://wp.m.163.com',
11. 'referer': 'https://wp.m.163.com/',
12. 'sec-ch-ua': '" Not A;Brand";v="99", "Chromium";v="90", "Google Chrome";v="90"',
13. 'sec-ch-ua-mobile': '?0',
14. 'sec-fetch-dest': 'empty',
15. 'sec-fetch-mode': 'cors',
16. 'sec-fetch-site': 'same-site',
17. 'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/90.0.4430.72 Safari/537.36'
18. }
20. url = 'https://c.m.163.com/ug/api/wuhan/app/data/list-total?t=324664221330'
22. **if** \_\_name\_\_ == "\_\_main\_\_":
23. session = requests.session()
24. session.cookies = CookieJar()
25. response = session.get(url=url, headers=headers1)
26. INfor = json.loads(response.text)
27. json = INfor['data']['chinaDayList']
28. **print**(json)
30. f = Workbook()
31. sheet = f.active
32. sheet.title = "Sheet1"
33. row0 = ["日期", "境外输入", "累计确诊", "累计死亡", "累计治愈"]
34. sheet["A1"].value = row0[0]
35. sheet["B1"].value = row0[1]
36. sheet["C1"].value = row0[2]
37. sheet["D1"].value = row0[3]
38. sheet["E1"].value = row0[4]
39. counter = 2
40. **for** i **in** range(len(json)-1):
41. json\_1=json[i]
42. sheet["A" + str(counter)].value = json\_1['date']                         #日期
43. sheet["B" + str(counter)].value = json\_1['total']['input']               #境外输入
44. sheet["C" + str(counter)].value = json\_1['total']['confirm']             #累计确诊
45. sheet["D" + str(counter)].value = json\_1['total']['dead']
46. sheet["E" + str(counter)].value = json\_1['total']['heal']
47. counter = counter + 1
48. f.save("Covid19.xlsx")
49. **print**("检索完毕，请查阅！")
50. input()

# 七、课设记录与小结·总结汇报交流

1、网络安全国策化

网络安全（Cyber Security）是指网络系统的硬件、软件及其系统中的数据受到保护，不因偶然的或者恶意的原因而遭受到破坏、更改、泄露，系统连续可靠正常地运行，网络服务不中断。

2020年4月27日，国家互联网信息办公室、国家发展和改革委员会、工业和信息化部、公安部、国家安全部、财政部、商务部、中国人民银行、国家市场监督管理总局、国家广播电视总局、国家保密局、国家密码管理局共12个部门联合发布《网络安全审查办法》，于2020年6月1日起实施。2021年2月，评选为2020年中国网络安全大事件。

2、社会生活信息化

社会生活信息化是以计算机信息处理技术和传输手段的广泛应用为基础和标志的新技术革命，影响和改造社会生活方式与管理方式的过程。它是指包括经济、科技、教育、军事、政务、日常生活等在内的整个社会体系不断采用先进的信息技术，建立各种信息网络，大力开发有关人们日常生活的信息资源，由此丰富人们精神生活，拓展人们活动时空的过程。当社会生活实现极大程度的信息化时，则意味我们已经进入了信息社会。

社会生活信息化指在经济生活全面信息化的进程中，人类社会生活的其他领域也逐步利用先进的信息技术，建立起各种信息网络;同时，大力开发有关人们日常生活内容，不断丰富人们的精神文化生活，提升生活质量的过程。社会信息化是信息化的高级阶段，它是指一切社会活动领域里实现全面的信息化。是以信息产业化和产业信息化为基础，以经济信息化为核心向人类社会活动的各个领域逐步扩展的过程，其最终结果是人类社会生活的全面信息化，主要表现为:信息成为社会活动的战略资源和重要财富，信息技术成为推动社会进步的主导技术，信息人员成为领导社会变革的中坚力量。

3、安全技术多元化

信息化的各个领域都存在安全，从软件到硬件再到协议，各个领域都需要安全技术的支撑。没有安全的领域就好像是裸奔，没有任何隐私可言，也很可能会造成各种经济损失。