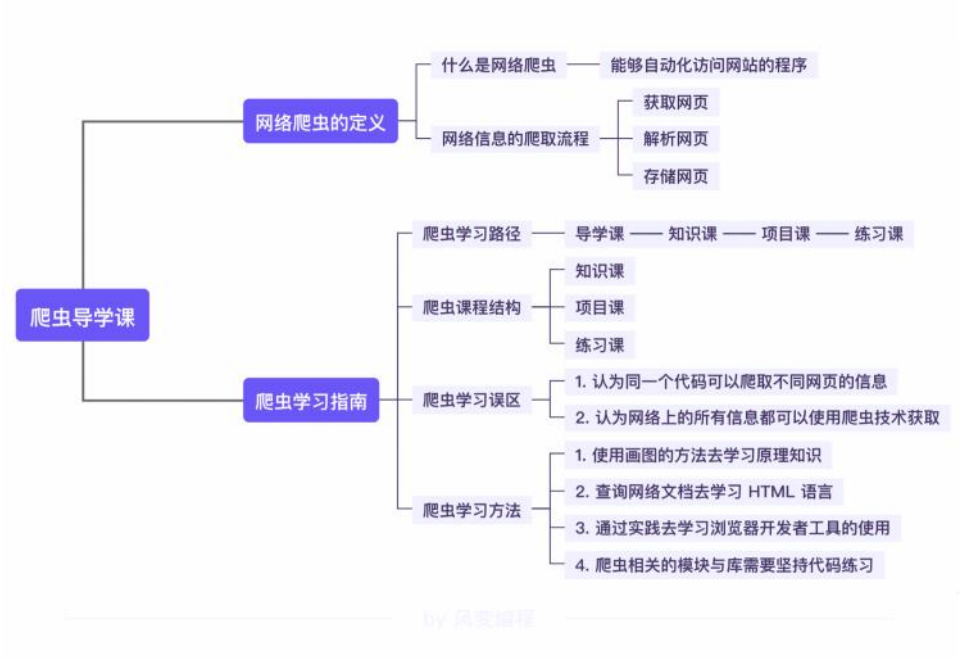
第二关 启蒙书苑：获取网页

2022年11月30日 10:03

1. 课前复习

在开始今天的课堂之前，我们先来复习一下上一节课的知识点。



上一节课我向你介绍了网络爬虫的定义，以及爬虫课程的学习指南。

下面就让我出一道题考察一下你的掌握程度。

单选题

以下选项中，哪一项展示了正确的爬虫流程？

A.
获取网页 -> 存储数据 -> 解析网页

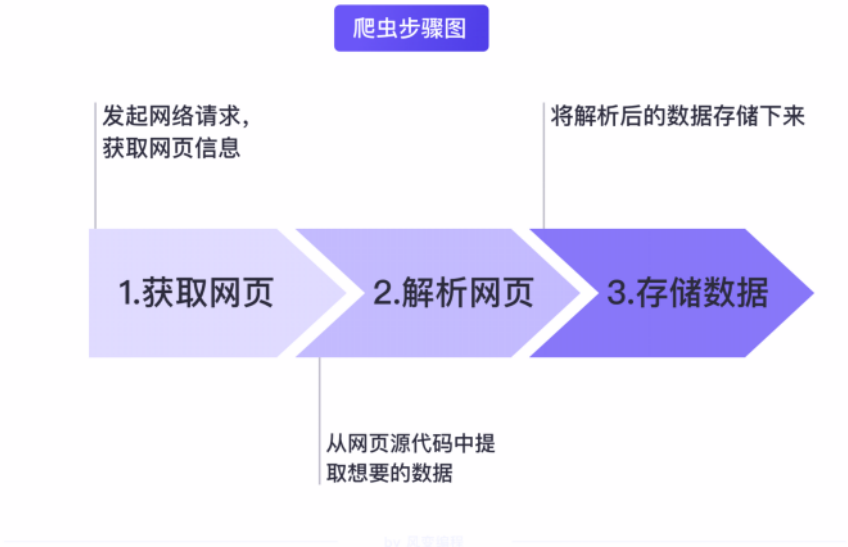
B.
获取网页 -> 解析网页 -> 存储数据

C.
获取网页 -> 下载网页 -> 解析网页 -> 存储数据

D.
获取网页 -> 下载网页 -> 存储数据 -> 解析网页

回答正确

答对了没有？没有的话就再看一遍这三个步骤的总结图吧！答对也可以再巩固一下。



2. 今天学什么？

这节课我先重点教你网络爬虫第一步：获取网页的相关知识，除此之外还准备向你传授一些网络爬虫的伦理。

由于今天这节课是你爬虫领域的第一节知识课，学习起来难免会遇到些问题。

在学习过程中，如果遇到了自己解决不了的问题请及时联系你的助教老师哦。

当然也别忘了导学课向你介绍的几个学习方法，下面就让我们开始今天的学习吧！

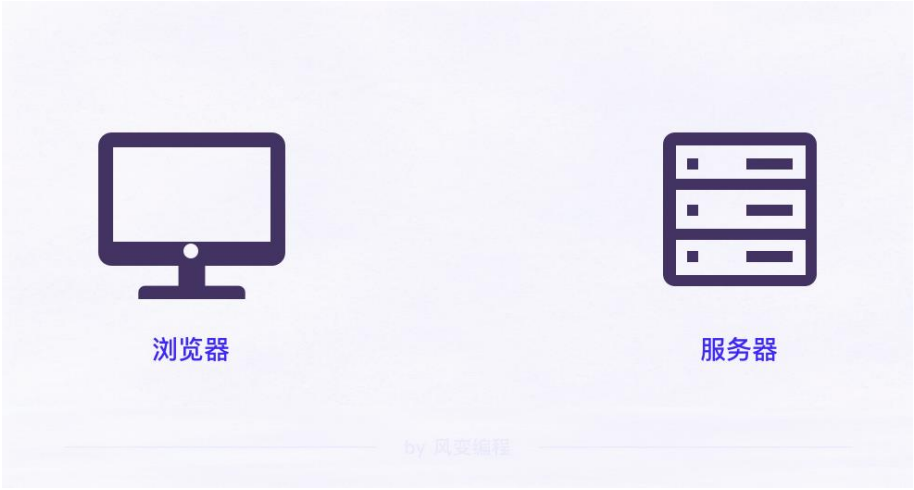
3. 获取网页

随着科技的普及，上网浏览信息已经成为了一项几乎人人都会的技能。但从输入网址，到网页信息加载完毕这一过程具体发生了什么，知道的人就没那么多了。

实际上，这一过程叫做网络请求。接下来我将带你详细了解网络请求的工作原理。这些内容，有助于你更进一步认识爬虫的基本原理。

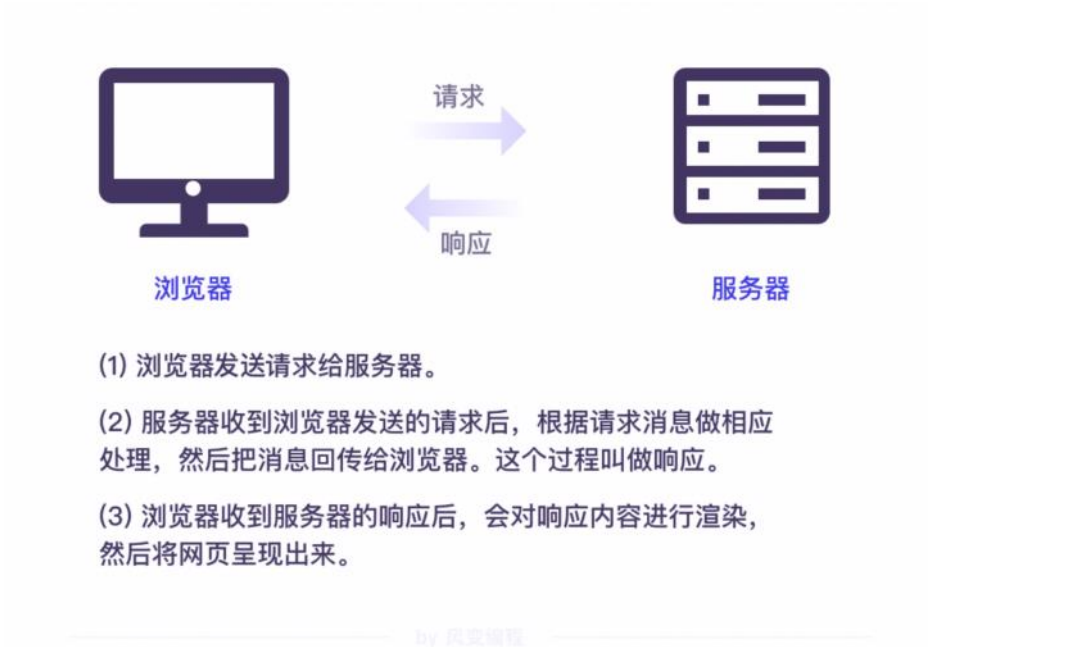
3.1 网络请求工作原理

当我们访问网站时，首先我们会在浏览器中输入网页链接，然后按回车。这个动作其实是浏览器向网页所在服务器发送了一个请求。



服务器接收到浏览器的请求后进行处理，返回响应内容，传给浏览器。浏览器再对响应内容进行渲染，将网页呈现了出来。

所以浏览器与服务器之间，有这么一层关系：先请求，后响应。



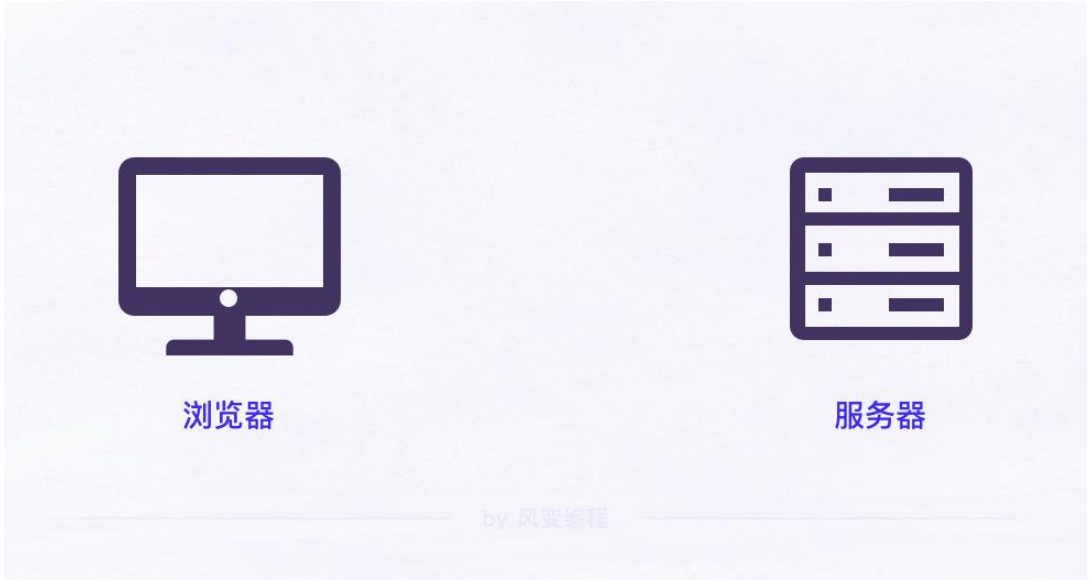
学习爬虫，最初的操作便是模拟浏览器向服务器发出请求。还记得我刚才介绍的爬虫第一步 —— 获取网页吗？这其中最关键的部分就是发起一个请求并发送给服务器。

3.2 requests.get()函数

Requests 库就是这样一种使用简洁的 Python 第三方库，能够帮我们发起各种网络请求。接下来就让我来教你使用 Requests 库发起网络请求。

Requests 库内置了很多函数来帮我们实现各种方法的网络请求，像 requests.get()就是 Requests 库中用来发起 GET 请求的函数。

GET 请求用于从服务器获取数据，是一种比较常用的请求方法，像我们平时在浏览器中直接输入网址回车，这便发起了一个 GET 请求。



当然，网络请求也有其他不同的方法，但我们的课程暂时还用不到，你可以看下面的图片了解一下。

网络请求常用方法	
GET	请求从服务器获取数据
POST	请求向服务器提交数据
PUT	请求向服务器更新数据
……	……

下面让我们来运行体验一下调用requests.get()函数发起网络请求。
我们这次先试着向百度首页：https://www.baidu.com/ 发起网络请求吧！
先看看下面这串代码。阅读完代码，点击运行即可。

```
# 仔细阅读完代码，点击运行即可。
import requests
# 发送请求，并把响应结果赋值在变量res上。
res = requests.get('https://www.baidu.com/')
```

#仔细阅读完代码，点击运行即可。
importrequests
#发送请求，并把响应结果赋值在变量res上。
res=requests.get('https://www.baidu.com/')

先导入 Requests 库，再调用 requests.get()函数，参数位置直接填写 URL 地址的字符串格式，把结果赋值到变量 res 上。
就这样，通过两行代码我们就成功发起了一个网络请求。
什么？你不知道URL是啥？
URL中文全称叫统一资源定位符，这个你不用记。你只要知道我们常说的**网址/链接就是一个 URL** 就行了。
它其中包含了协议、主机域名和路径。通过这样一个链接我们就能在互联网上找到这个资源。

上面图片是我们官网的 URL，开头的 https 是协议，我马上就会讲；双斜杠// 后的 wp.forchange.cn 是主机域名，你可以把它理解为服务器在网络上的位置。
后面的 /resources/ 是路径，就跟你之前见过的系统路径一样，表示资源在主机域名下的位置，各级目录会用 / 隔开。

其实系统路径跟 URL 一样，本质都是用来定位资源位置的。只不过一个是你的系统本地的位置，另一个是互联网中的位置。
那么这个开头的 https 是什么来路呢?
http与https
我们日常浏览的网页 URL 通常都会看到 http 或 https，这些都是访问资源需要的协议类型。
除此之外，也会有像 ftp、sftp、smb 开头的 URL，它们也都是协议类型，但我们这个系列课程里不会出现。
http 是一种从网络传输数据到本地浏览器的传送协议，我们一般也称它为 http 协议。
它基于“请求与响应”模式，能保证高效而准确地传送超文本文档。我们一般情况下说的网络请求也可以被称作 http 请求。
有同学会问了，那为什么有的请求是 http，有的请求是 https? 从英文角度来说，好像给 http 协议加了复数形式? 但这里的 https 可不是复数形式哦。
在 http 协议中，数据没有经过加密处理，所以你在网络上输入的账号密码，个人信息等数据如果被人获取的话就非常危险。
而 https 协议会在 http 协议的基础上对数据进行加密处理，相对于 http 来说更加安全、可靠。
所以现在市面上的主流网站几乎都采用的是 https 协议。这些网络协议知识是深入网络爬虫技术的基础。
目前你只需要了解 http 的传输是基于请求与响应就够了。说到这里，还记得我刚才讲过的服务器接收到浏览器的请求后会返回响应信息吗?
使用 requests 库发起请求后，我们如何查看响应内容呢? 我们试着直接打印 requests.get() 的函数结果吧。

```
1  # 仔细阅读完代码，点击运行即可。
2  import requests
3  # 发送请求，并把响应结果赋值在变量res上
4  res = requests.get('https://www.baidu.com/')
5  print(res)
6
```

```
问题  输出  调试控制台  终端  JUPYTER

Windows PowerShell
版权所有 (C)  Microsoft Corporation。保留所有权利。

安装最新的 PowerShell，了解新功能和改进! https://aka.ms/PSWindows

PS D:\PythonTest> & C:/Users/heaven/AppData/Local/Programs/Python/Python311/python.exe d:/PythonTes
t/python体验第0关.py
<Response [200]>
PS D:\PythonTest>
```

3.3 响应

我们不妨使用 type() 函数来查看一下 requests.get() 函数返回结果的类型。
type()函数你肯定不陌生，这个代码你自己写写看吧。

```
1  # 仔细阅读完代码，点击运行即可。
2  import requests
3  # 发送请求，并把响应结果赋值在变量res上
4  res = requests.get('https://www.baidu.com/')
5  print(res)
6  print(type(res))
7
```

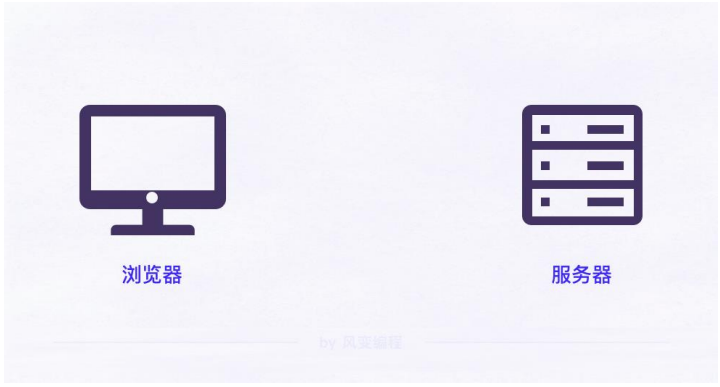
```
问题  输出  调试控制台  终端  JUPYTER

Windows PowerShell
版权所有 (C)  Microsoft Corporation。保留所有权利。

安装最新的 PowerShell，了解新功能和改进! https://aka.ms/PSWindows

PS D:\PythonTest> & C:/Users/heaven/AppData/Local/Programs/Python/Python311/python.exe d:/
<Response [200]>
<class 'requests.models.Response'>
PS D:\PythonTest>
```

这里我们可以看到，requests.get() 函数返回结果是一个属于 requests.models.Response 类的对象。
Response 就是英文里响应的意思，Response 对象顾名思义，就是一个包含各种网络请求响应信息的对象。
一般来说，服务器返回给浏览器的响应内容，主要包含了响应状态码、响应头和响应体等信息。
响应头的信息现阶段你还用不到，这次就先不跟你介绍了，我们重点来看响应状态码和响应体信息。
当服务器收到网络请求时，会返回一个三位数字的代码响应浏览器的请求，表示服务器对于这个请求的响应状态，我们称之为响应状态码。
其中200代表服务器成功处理了请求。当然也有很多其它的响应状态码，比如你经常会见到的404指的就是服务器无法根据请求找到资源。



一般来说状态码中第一个数字就定义了状态码的类型，下面有一个表格，供你参考不同的状态码代表什么，但不需要记住它们，在遇到问题的时候查询就好。

常见响应状态码解释

响应状态码	说明	举例	说明
1xx	请求收到	100	继续提出请求
2xx	请求成功	200	成功
3xx	重新定向	305	应使用代理访问
4xx	客户端错误	404	无法提供信息
5xx	服务器错误	503	服务不可用

3.3.1 Response对象 —— status_code属性

这里具体的响应信息我们可以通过调用 Response 对象中的属性去获得。

像 Response.status_code 属性，里面就是响应状态码。

先看看下面这串代码。阅读完代码，点击运行即可。

```
# 仔细阅读完代码，点击运行即可。
import requests
# 发送请求，并把响应结果赋值在变量res上
res = requests.get('https://www.baidu.com/')
print(res)
print(type(res))
print(res.status_code)
```

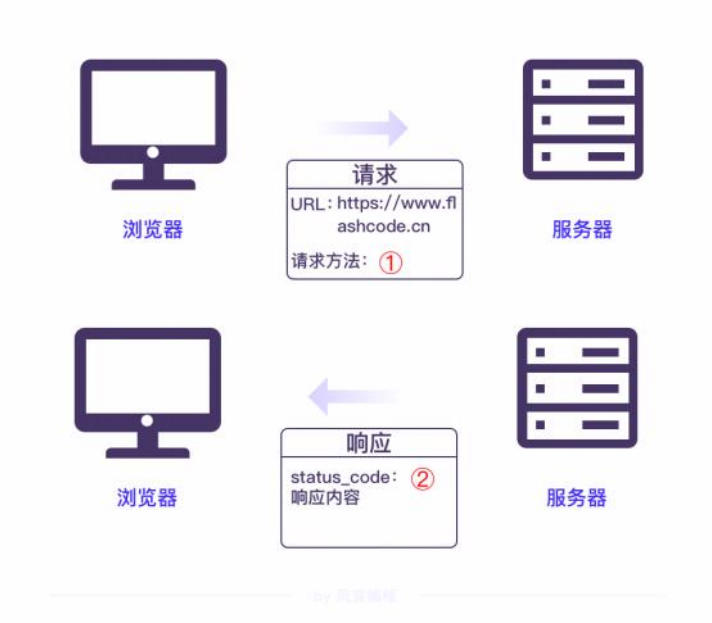
```
问题  输出  调试控制台  终端  JUPYTER

Windows PowerShell
版权所有 (C) Microsoft Corporation。保留所有权利。

安装最新的 PowerShell，了解新功能和改进！ https://aka.ms/PSWindows

PS D:\PythonTest> & C:/Users/heaven/AppData/Local/Programs/Python/Python311/python.exe d:/Pyt
<Response [200]>
<class 'requests.models.Response'>
PS D:\PythonTest> & C:/Users/heaven/AppData/Local/Programs/Python/Python311/python.exe d:/Pyt
<Response [200]>
<class 'requests.models.Response'>
200
PS D:\PythonTest>
```

下面我来出一道练习考察你对目前学到的网络请求知识的理解。先来看张图：



单选题

上图中，浏览器向服务器发起了一个获取服务器数据的请求，服务器成功接收请求并响应。那么①位置的请求方法，②位置的状态码应该分别是什么？

A.
① POST；② 200

B.
① GET；② 404

C.
① POST；② 404

D.
① GET；② 200

分析

答案是 D

访问页面的请求方法为 GET，服务器成功接收请求并响应的状态码为 200。POST 方法我还没有教你，404 状态码为页面未找到。

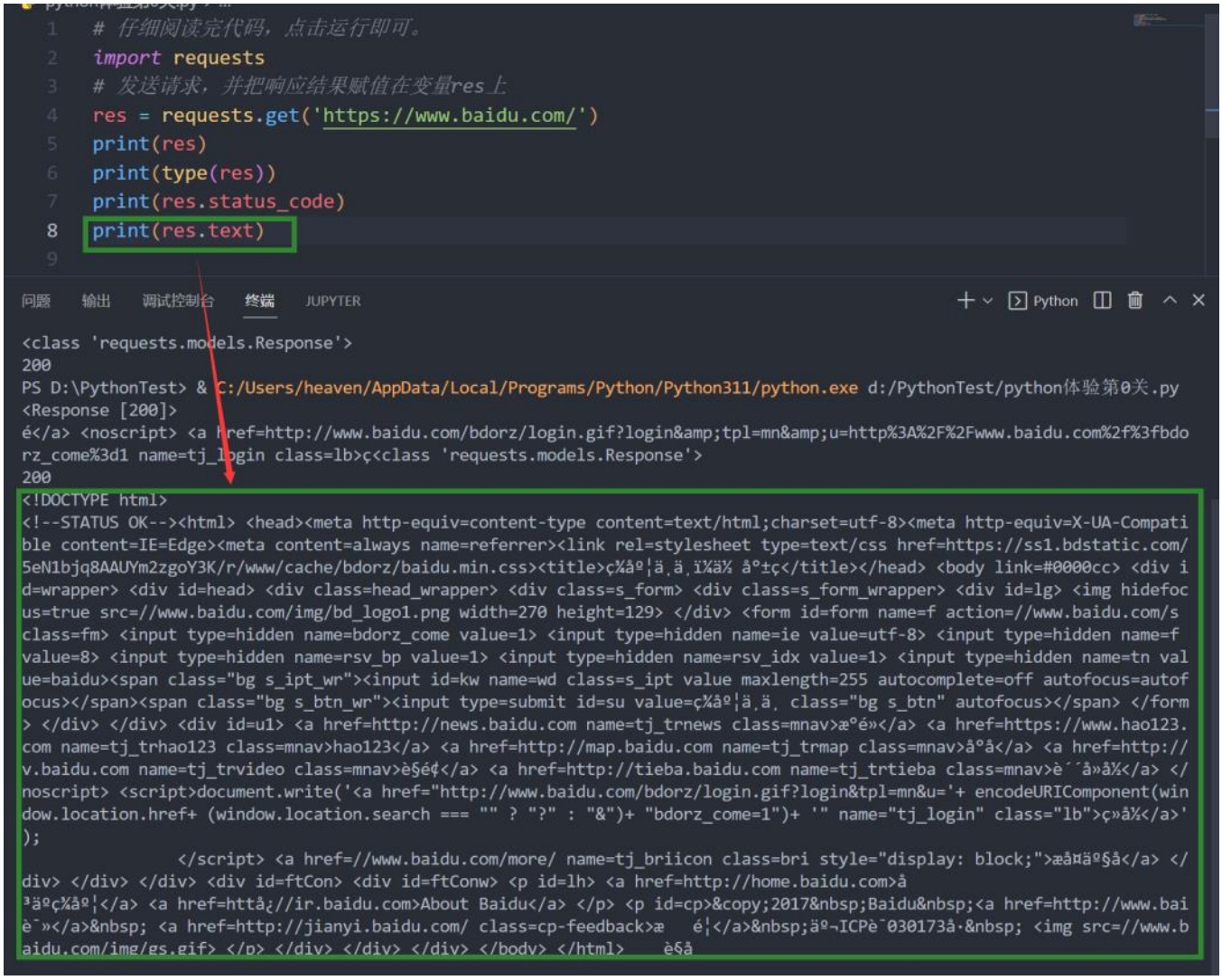
3.3.2 Response对象 —— text属性

通过响应状态码判断网络请求成功后，我们就可以查看具体的响应内容了。

响应的正文数据都在响应体中。我们请求网页时，它的响应体就是网页的 HTML 代码。浏览器会将其直接渲染成相应的页面展示出来。

在爬虫程序中，我们可以通过调用 Response.text 属性，查看具体的响应内容。

看看下面这串代码。阅读完代码，点击运行即可。



我们可以看到 Response.text 可以得到网页代码的字符串形式，现在的你看这些 HTML 代码可能会一头雾水。没关系，下一节课我就会教你查看并修改网页的 HTML 源码。

requests库默认的字符编码是英文的。我们这次获取的是中文网页，所以字符编码格式也要设置为中文的。

3.3.3 Response对象 —— encoding属性

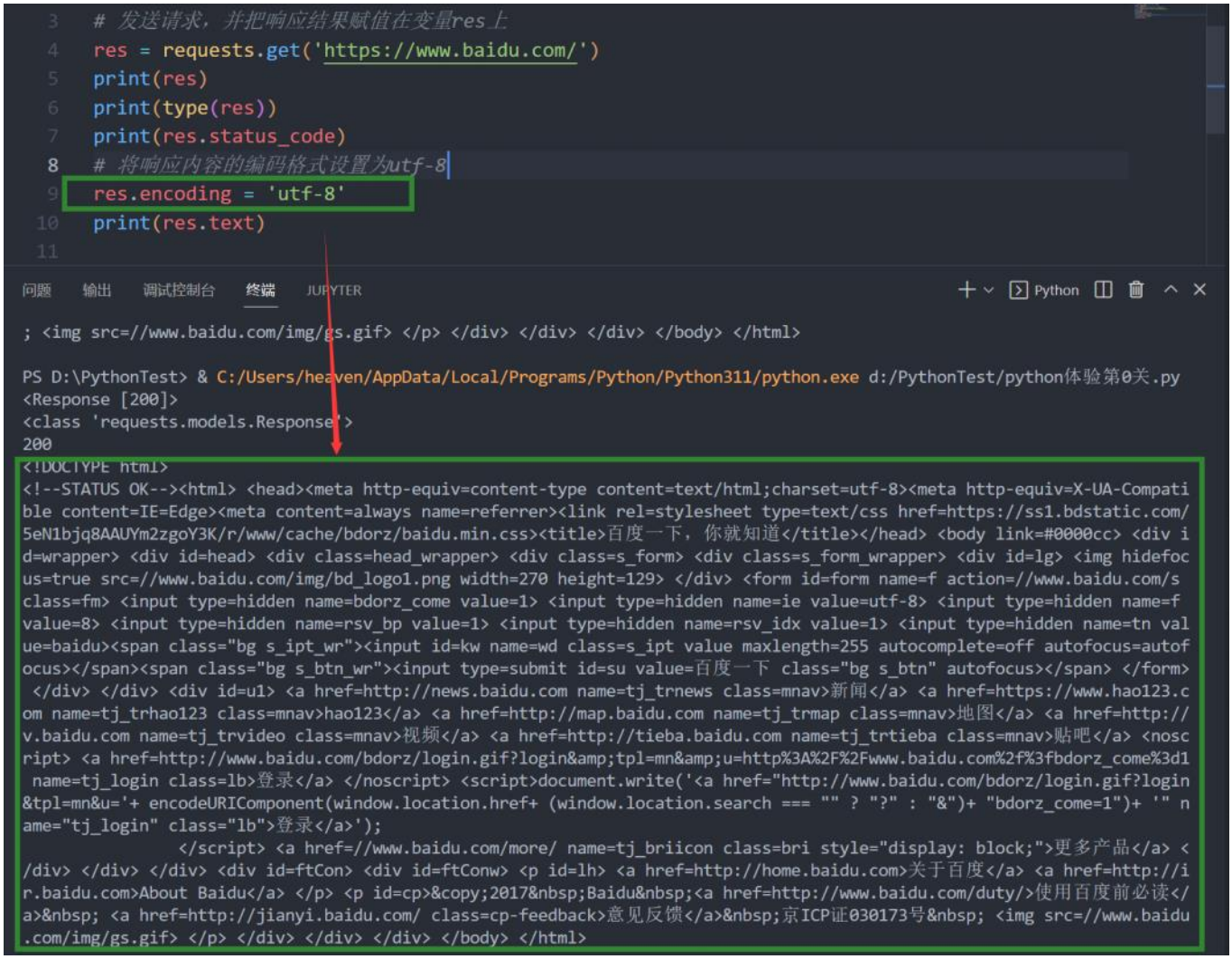
我们只需将Response对象的encoding属性设置为中文字符的编码格式，响应内容就可以正常显示了。

常用的中文字符编码格式有utf-8以及GBK，这次时间有限，编码这部分我就不展开讲了。

看看下面这串代码。阅读完代码，点击运行即可。

刚才我有说过，常用的中文字符编码格式有 utf-8 以及 GBK。如果GBK不行就再试试将 encoding 属性设置为 utf-8。

下面你自己来试一下吧，去将响应内容的编码格式设置为utf-8吧！要记得加引号噢。



现在的主流网站几乎都是用 utf-8 作为中文字符的编码格式，所以你之后再使用爬虫时要记住，响应成功后将encoding 属性先设置为 utf-8。
这样就能得到大部分中文网站正确编码格式下的响应内容。当然，如果不行就再试一下 GBK。
市面上主流网站的中文字符无外乎就这两种编码，使用枚举法挨个去试就好了。

4. 网络爬虫伦理

4.1 网络爬虫引发的问题

就像是两个人在来来往往的相处中，会考虑对方的感受；在互联网的世界中，我们也要考虑一下服务器对爬虫的感受是怎样的。
通常情况下，服务器不太会在意小爬虫。但是，服务器会拒绝频率很高的大型爬虫和恶意爬虫，因为这会给服务器带来极大的压力或伤害。
当然，有些时候没有恶意但质量糟糕的爬虫，也会导致服务器或者路由器瘫痪。个人爬虫，如果过多的人使用，可能占用过多服务器资源。
除此之外，服务器上的数据有产权归属，网络爬虫获取数据如果用来牟利也会带来法律风险。
《弟子规》有云:“用人物，须明求，倘不问，即为偷”。
在互联网世界中，存在着一个 Robots 协议，全称网络爬虫排除标准（Robots exclusion protocol）。是网站开发者用来告知网络爬虫哪些页面可以抓取，哪些不行的协议。
Robots协议是互联网爬虫的一项公认的道德规范，这个协议用来告诉爬虫，哪些页面是可以抓取的，哪些不可以。
如何查看网站的 Robots 协议呢？很简单，在网站的主机域名后加上/robots.txt就可以了。
比如我们要查看闪光读书网站的 Robots 协议的话，就直接访问 https://wp.forchange.cn/robots.txt 查看就好了。
下面我来以闪光读书网站的 Robots 协议为例，教你如何查看 Robots 协议。
查看 Robots 协议

```
User-agent: Baiduspider
Disallow: /

User-agent: *
Disallow: /
Allow: /resources/
Allow: /psychology/
Allow: /economic/
Allow: /history/
Allow: /biography/
Allow: /food/
Allow: /cartoon/
```

在 Robots 协议中，User-agent代表的是爬虫身份。
比如User-agent: Baiduspider代表的就是百度的爬虫，User-agent: *代表的是未指定身份的所有爬虫，像我们就属于User-agent: *。
Robots 协议对于不同的爬虫身份有着不一样的限制。
Disallow:声明了网站禁止百度爬虫爬取的内容。冒号后面代表限制爬虫的路径。
Disallow:/表示闪光读书禁止百度爬取https://wp.forchange.cn/下的所有内容。
不过这些我们不用担心，禁止只是百度爬虫。我们只需要查看下面User-agent: *的限制。
首先我们看到一个跟百度爬虫一样的Disallow: /，这里并不意味着我们对于闪光读书网站的爬取受到了限制。
因为往后看，我们还能看到一些Allow的路径。
在Robots协议中，Allow：冒号后面表示的是允许爬取的路径。
这几行协议表示主机域名https://wp.forchange.cn下的/resources/，/psychology/…/cartoon/路径中的所有内容都是可以爬取的。
Allow与Disallow组合使用，表示闪光读书网站除了Allow允许爬取的几个路径外，其它路径下的内容都不可爬取。
我们这次项目需要爬取的书籍信息也基本都在这几个路径下，所以我们可以放心大胆的去爬。

了解完这些我们再来查看一下其他网站的 Robots 协议。就拿你想爬的淘宝网为例吧。
它的 Robots 协议 URL 为：https://www.taobao.com/robots.txt
点开后我们可以看到：

```
User-agent: Baiduspider
Disallow: /

User-agent: baiduspider
Disallow: /
```

我们可以从名字推断出来，淘宝网的 Robots 协议限制两个百度爬虫：Baiduspider，baiduspider爬取淘宝网的所有内容。
所以，当你在百度搜索“淘宝网”时，会看到下图的这两行小字。

因为百度很好地遵守了淘宝网的robots.txt协议。自然，你在百度中也查不到淘宝网的具体商品信息了。
像我们爬虫的话主要是看星号部分User-Agent: *，这是不是意味着我们就能随意爬取淘宝网的商品信息了呢?
假设我们要爬取淘宝网中所有衬衫的商品信息。

从分类中点进去我们会发现
：

主机域名已经变成了 https://s.taobao.com/ ，我们再来查看 https://s.taobao.com/robots.txt 下的 Robots 协议。

```
User-agent: *
Disallow: /
```

简单粗暴，禁止所有爬虫爬取任何信息。不过也可以理解，毕竟淘宝网访问量那么大，如果不对爬虫加以限制，服务器资源会被消耗的非常严重。
虽然Robots协议只是一个道德规范，和爬虫相关的法律也还在建立和完善之中，但我们在爬取网络信息时，应该有意识地去遵守这个协议。
网站的服务器被爬虫爬得多了，也会受到较大的压力，因此，各大网站也会做一些反爬虫的措施。不过呢，有反爬虫，也就有相应的反反爬虫。
爬虫就像是核技术，人们可以利用它去做有用的事，也能利用它去搞破坏。
恶意消耗别人的服务器资源，是一件不道德的事，恶意爬取一些不被允许的数据，还可能会引起严重的法律后果。
工具在你手中，如何利用它是你的选择。当你在爬取网站数据的时候，别忘了先看看网站的Robots协议是否允许你去爬取。维持良好的互联网秩序，是我们该做的事。
了解完这些下面就让我考察一下你对爬虫伦理的认识。
以下是闪光科技官网的 Robots 协议。

```
User-Agent:  *
Disallow:  /users/
```

单选题

下面选项中哪个选项是符合爬虫伦理的？

A.
爬取页面 https://prod.pandateacher.com/web-simulation/shining-technology/about.html 里的信息

B.
爬取 https://prod.pandateacher.com/users/ 路径下的内容

C.
将页面 https://prod.pandateacher.com/web-simulation/shining-technology/about.html 里的内容发布到微信公众号收获打赏

D.
将发送请求的代码写进一个死循环（while True:）后执行
分析
答案是 A

在这个 Robots 协议中，只有/users/路径是不允许爬虫爬取信息的。
错了，闪光科技官网只允许爬虫访问/shining-technology/。

5. 总结与练习

以上，就是本关的学习内容。下面我们来总结一下这节课所学的知识。

5.1 知识总结

总结一下今天的内容，来看下图：

下面我们来做一道综合练习来复习一下这节课的内容吧。

5.2 课后练习

打印网页响应内容

公司网站上周刚更新了一遍，里面有我们闪光科技 CEO - WF 的献词，URL是：
https://prod.pandateacher.com/web-simulation/shining-technology/about.html
【温馨提示】你可以复制这里的网址到浏览器中打开，然后对照着课堂内容来学习。
请你结合所学知识，根据代码注释在右侧填充代码，点击继续后开始。

```
import requests
# 结合所学知识，根据代码注释填充代码。
# 导入requests库


# 向闪光科技献词网页的URL发送网络请求，并把响应结果赋值在变量res上
res = requests.get('https://prod.pandateacher.com/web-simulation/shining-technology/about.html')

# 打印响应的状态码

# 使用枚举法先将响应内容的编码格式设置为utf-8

# 调用Response对象的text属性，获取响应内容

# 打印响应内容
```

答案

```python
# 仔细阅读完代码，点击运行即可。
import requests
# 结合所学知识，根据代码注释填充代码。
# 导入requests库
import requests
# 向闪光科技献词网页的URL发送网络请求，并把响应结果赋值在变量res上
res = requests.get('https://prod.pandateacher.com/web-simulation/shining-technology/about.html')
# 打印响应的状态码
print(res.status_code)
# 使用枚举法先将响应内容的编码格式设置为utf-8
res.encoding = 'utf-8'

# 调用Response对象的text属性，获取响应内容
print(res.text)

# 打印响应内容
```

今天的课程到这里就结束了。内容不多，但大部分都是新知识，回去可以试着发起几个网站请求练习一下。
HTML 语言的结构层级关系十分严谨，了解后我们就可以很轻易的在网页代码中提取我们想要的数据。
具体的数据提取操作我之后就会教给你，敬请期待吧！

第二关 第二关练习

2022年11月30日 12:13

1. 知识回顾

本节课为爬虫第二关的练习课。在正式开始练习前，让我们通过下面的思维导图回顾一下第二关的重要知识点：

为了让你能更好的巩固第二关的知识点，做好学习下一关的准备。请你试着完成下方的练习吧：

2. 牛刀小试

选择题 1 - 网络信息的爬取流程

单选题

以下选项中，哪一项展示了正确的爬虫流程？

A.
解析网页 -> 获取网页 -> 存储数据

B.
获取网页 -> 下载网页 -> 解析网页

C.
获取网页 -> 下载网页 -> 存储数据

D.
获取网页 -> 解析网页 -> 存储数据

回答正确

网络爬虫的流程主要可以分为三步，分别是：**获取网页**、**解析网页**以及**存储数据**。

本题的知识包含 网络信息的爬取流程，如果有所遗忘，课后请适当进行回顾复习。

by 风变编程

选择题 2 - 统一资源定位符

已知书籍《全球通史》的网址为： https://wp.forchange.cn/history/6704/ **。**

单选题

请问，上方网址的**协议**、**主机域名**和**路径**，分别为哪部分？

A.
https；wp.forchange.cn；/history/6704/

B.
https；wp.forchange.cn/history；/6704/

C.
https://wp.forchange.cn；/history；/6704/

D.
https；wp.forchange.cn/history；/6704/

分析

答案是 A

（1）https 是协议；
（2）双斜杠 // 后的 wp.forchange.cn 是主机域名，可以理解为服务器在网络上的位置；
（3）/history/6704/ 是路径，表示资源在主机域名下的位置，各级目录会用 / 隔开。

本题的知识包含 统一资源定位符（URL），**如果有所遗忘，课后请适当进行回顾复习。**

选择题 3 - Robots 协议

单选题

想要查看某个网站的 Robots 协议，需要在网站的主机域名后加上以下哪个选项的内容？

A.
/robots.text

B.
/robot.txt

C.
/robots.txt

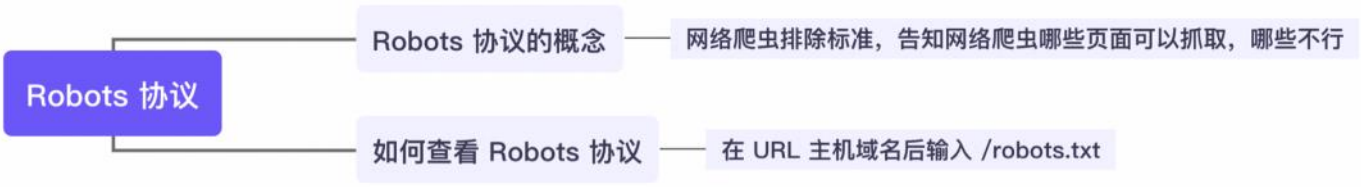
D.
/Robots.txt

分析

答案是 C

在网站的主机域名后加上 /robots.txt，可以查看网站的 Robots 协议。

本题的知识包含 Robots 协议，如果有所遗忘，课后请适当进行回顾复习。



选择题 4 - 爬虫伦理

以下是《今日头条》网站的部分 Robots 协议：

```
User-agent: *
Disallow: /
Allow: /complain/
Allow: /media_partners/
Allow: /about/
Allow: /user_agreement/
Allow: /$
```

多选题

下面哪几个选项是不符合爬虫伦理的？

A.
爬取页面 https://www.toutiao.com/friend_link/ 里的信息

B.
频繁地爬取页面 https://www.toutiao.com/user_agreement/ 里的信息

C.
随意爬取《今日头条》网站的所有信息，并用以商业盈利

D.
爬取页面 https://www.toutiao.com/about/ 里的信息

分析

答案是 A、B、C

选项 A：该网站的 Robots 协议中，Allow 允许的路径中没有提到 /friend_link/，因此不允许爬取该路径的信息；
选项 B：过于频繁地爬取网站，会给服务器产生巨大的压力，网站可能封锁 IP，甚至采取进一步的法律行动；
选项 C：服务器上的数据有产权归属，网络爬虫获取数据如果用来牟利会带来法律风险。

本题的知识包含 爬虫伦理，**如果有所遗忘，课后请适当进行回顾复习。**

选择题 5 - 响应状态码

单选题

服务器成功接收请求并响应的状态码为？页面未找到而导致请求失败的状态为？

A.
200；400

B.
200；404

C.
400；200

D.
204；400
分析
服务器成功接收请求并响应，会返回状态码 200；
页面未找到而导致请求失败，会返回状态码 404。
本题的知识包含 常用响应状态码，如果有所遗忘，课后请适当进行回顾复习。

常见响应状态码解释			
响应状态码	说明	举例	说明
1xx	请求收到	100	继续提出请求
2xx	请求成功	200	成功
3xx	重新定向	305	应使用代理访问
4xx	客户端错误	404	无法提供信息
5xx	服务器错误	503	服务不可用

3. 代码实战
实战 1 - 网页请求
网页请求
练习介绍
已知书籍《全球通史》的网址为：https://wp.forchange.cn/history/6704/ 。
题目要求
请你补充代码，实现以下功能：
1）使用 Requests 库中的 GET 请求，向该网站发起网络请求；
2）打印服务器返回的响应状态码。

```
# 导入 requests 库


# 使用 Get 请求向网站发起请求


# 打印服务器返回的响应状态码
```

答案

```python
# 导入 requests 库

import requests
# 使用 Get 请求向网站发起请求
res = requests.get(' https://wp.forchange.cn/history/6704/ ')

# 打印服务器返回的响应状态码

print(res.status_code)
```

实战 2 - 打印网页响应内容

打印网页响应内容

练习介绍

已知《电影天堂》的网址： https://www.dytt8.net/index2.htm 。

题目要求

请你补充代码，实现以下功能：

1）将请求该网页返回的 Response 对象的编码格式设置为 gbk；

2）获取网页的响应内容。

```python
# 导入 requests 库


# 使用 Get 请求向网站发起请求


# 将 Response 对象的编码格式设置为 gbk


# 打印响应内容
```

答案

```python
# 导入 requests 库

import requests

# 使用 Get 请求向网站发起请求
res = requests.get('https://www.dytt8.net/index2.htm ')

# 将 Response 对象的编码格式设置为 gbk
res.encoding = 'gbk'

# 打印响应内容
print(res.text)
```

恭喜你完成本节课的练习题。