

开发过程文档

问题 1

1、代码初步完成后，运行 webserver，然后访问网页，网页可正常访问，但是在正常的信息中会包括 list index out of range 的错误。

```
Exception in thread Thread-3:  
Traceback (most recent call last):  
  File "C:\Users\admin\AppData\Local\Programs\Python\Python36\lib\threading.py", line 916,  
    self.run()  
  File "C:\Users\admin\AppData\Local\Programs\Python\Python36\lib\threading.py", line 864,  
    self._target(*self._args, **self._kwargs)  
  File "D:/资料/webserver/src/webserver4.py", line 21, in service_client  
    ret = re.match(r"^[^/]+(/[^\ ]*)", request_lines[0])  
IndexError: list index out of range
```

2、上网搜索，发现出现这个错误信息的原因有两种情况：

第 1 种可能情况：list[index]index 超出范围

第 2 种可能情况：list 是一个空的 没有一个元素 进行 list[0]就会出现该错误

根据报错信息提示，可以判断问题出现在 request_lines[0]这一参数中，推测是 request_lines 为空，因此出现错误。由于 request_lines 是对 request 执行 splitlines 方法得到的，因此添加两个打印，分别打印 request 和 request_lines[0]。

```
request = new_socket.recv(1024).decode("utf-8")  
print("request:"+request)  
  
request_lines = request.splitlines()  
print("request_lines:"+request_lines[0])
```

3、修改后运行 webserver，访问网页

正确信息如下：

```

request:GET /index.html HTTP/1.1
Host: localhost:7890
Connection: keep-alive
sec-ch-ua: " Not;A Brand";v="99", "Microsoft Edge";v="91", "Chromium";v="91"
sec-ch-ua-mobile: ?0
Upgrade-Insecure-Requests: 1
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,image/apng,*/*;q=0
Sec-Fetch-Site: same-origin
Sec-Fetch-Mode: navigate
Sec-Fetch-User: ?1
Sec-Fetch-Dest: document
Referer: http://localhost:7890/success.html
Accept-Encoding: gzip, deflate, br
Accept-Language: zh-CN,zh;q=0.9,en;q=0.8,en-GB;q=0.7,en-US;q=0.6
Cookie: Idea-47a386c3=0b9140fd-7397-4a47-92e5-4b1948e71c4b

request_lines:GET /index.html HTTP/1.1
<_sre.SRE_Match object; span=(0, 15), match='GET /index.html'>
/index.html
HTTP/1.1 200 OK

```

出现错误信息:

```

Exception in thread Thread-4:
Traceback (most recent call last):
  File "C:\Users\admin\AppData\Local\Programs\Python\Python36\lib\threading.py", line 916,
    self.run()
  File "C:\Users\admin\AppData\Local\Programs\Python\Python36\lib\threading.py", line 864,
    self._target(*self._args, **self._kwargs)
  File "D:/资料/webserver/src/webserver4.py", line 15, in service_client
    print("request_lines:"+request_lines[0])
IndexError: list index out of range

request:

```

可以看到,此时 request 为空, 确认是 request_lines 为空导致的错误。

4、解决方案, 在提取 request 中文件名信息时, 先判断 response 是否为空, 不为空则进行, 为空则打印-----request not exist-----

```

request = new_socket.recv(1024).decode("utf-8")
print("request:" + request)

request_lines = request.splitlines(0)
response = ""

if request_lines:
    # GET /index.html
    # get post put del
    print("request_lines:" + request_lines[0])
    ret = re.match(r"[^/]+(?:/[^ ]*)", request_lines[0])

```

可以看出 request 为空时, 没有报错, 提示-----request not exist-----

```
request:
-----request not exist-----
```

Request 为空原因分析：在多线程当中，多个线程同时工作，只剩下一个请求的时候，第一个进入 `new_socket.recv(1024)` 函数的线程准备取走最后一个 request 时，后来的线程也在进行判断，当请求被取走了后，后来的线程在获取请求时就会产生 `#list index out of range` 这个报错。

```
Exception in thread Thread-1:
Traceback (most recent call last):
  File "C:\Users\admin\AppData\Local\Programs\Python\Python36\lib\threading.py", line 916,
    self.run()
  File "C:\Users\admin\AppData\Local\Programs\Python\Python36\lib\threading.py", line 864,
    self._target(*self._args, **self._kwargs)
  File "D:/资料/webserver/src/webserver4.py", line 66, in service_client
    print("body:" + html_content)
TypeError: must be str, not bytes
```

问题 2

1、访问 `fail.html` 文件时，文件中图片不显示。



2、后台出现报错信息

```
Exception in thread Thread-22:
Traceback (most recent call last):
  File "C:\Users\admin\AppData\Local\Programs\Python\Python36\lib\threading.py", line 916,
    self.run()
  File "C:\Users\admin\AppData\Local\Programs\Python\Python36\lib\threading.py", line 864,
    self._target(*self._args, **self._kwargs)
  File "D:/资料/webserver/src/webserver4.py", line 11, in service_client
    request = new_socket.recv(1024).decode("utf-8")
UnicodeDecodeError: 'utf-8' codec can't decode byte 0xfc in position 8: invalid start byte

Exception in thread Thread-23:
Traceback (most recent call last):
  File "C:\Users\admin\AppData\Local\Programs\Python\Python36\lib\threading.py", line 916,
    self.run()
  File "C:\Users\admin\AppData\Local\Programs\Python\Python36\lib\threading.py", line 864,
    self._target(*self._args, **self._kwargs)
  File "D:/资料/webserver/src/webserver4.py", line 11, in service_client
    request = new_socket.recv(1024).decode("utf-8")
```

3、考虑到图片不显示，检查 html 文件中图片显示代码

```
src="https://localhost:7890/404-not-found.jpg" width="400px"/></div>
```

检查图片是否有问题，替换为 index.html 中能访问的图片，flower.html，依然不显示图片

直接复制 index.html 中代码，可以访问

```
></div>
```

仔细观察，区别在于 https 和 http,将 fail.html 中 https 修改为 http,图片正常显示

```
img src="http://localhost:7890/404-not-found.jpg" width="400px"/>
```

localhost:63342/webserver/html/fail.html?_ijt=o9ndh3mmtrrh090n5du5k6fe9m

所访问的路径不存在!



[返回主页面](#)

后台请求和回应正常显示

```
request:GET /404-not-found.jpg HTTP/1.1
Host: localhost:7890
Connection: keep-alive
sec-ch-ua: " Not;A Brand";v="99", "Microsoft Edge";v="91", "Chromium";v="91"
sec-ch-ua-mobile: ?0
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.114 Safari/537.36 Edg/91.0.864.59
Accept: image/webp,image/apng,image/svg+xml,image/*,*/*;q=0.8
Sec-Fetch-Site: same-site
Sec-Fetch-Mode: no-cors
Sec-Fetch-Dest: image
Referer: http://localhost:63342/
Accept-Encoding: gzip, deflate, br
Accept-Language: zh-CN,zh;q=0.9,en;q=0.8,en-GB;q=0.7,en-US;q=0.6
Cookie: Idea-47a386c3=0b9140fd-7397-4a47-92e5-4b1948e71c4b

<socket.socket fd=616, family=AddressFamily.AF_INET, type=SocketKind.SOCK_STREAM, proto=0, laddr=('127.0.0.1', 7890), raddr=('127.0.0.1', 59260)>
```

http,https 分析:

HTTP（HyperText Transfer Protocol: 超文本传输协议）是一种用于分布式、协作式和超媒体信息系统的应用层协议。简单来说就是一种发布和接收 HTML 页面的方法，被用于在 Web 浏览器和网站服务器之间传递信息。

HTTP 默认工作在 TCP 协议 80 端口，用户访问网站 http:// 打头的都是标准 HTTP 服务。HTTP 协议以明文方式发送内容，不提供任何方式的数据加密，如果攻击者截取了 Web 浏览器和网站服务器之间的传输报文，就可以直接读懂其中的信息，因此，HTTP 协议不适合传输一些敏感信息，比如：信用卡号、密码等支付信息。

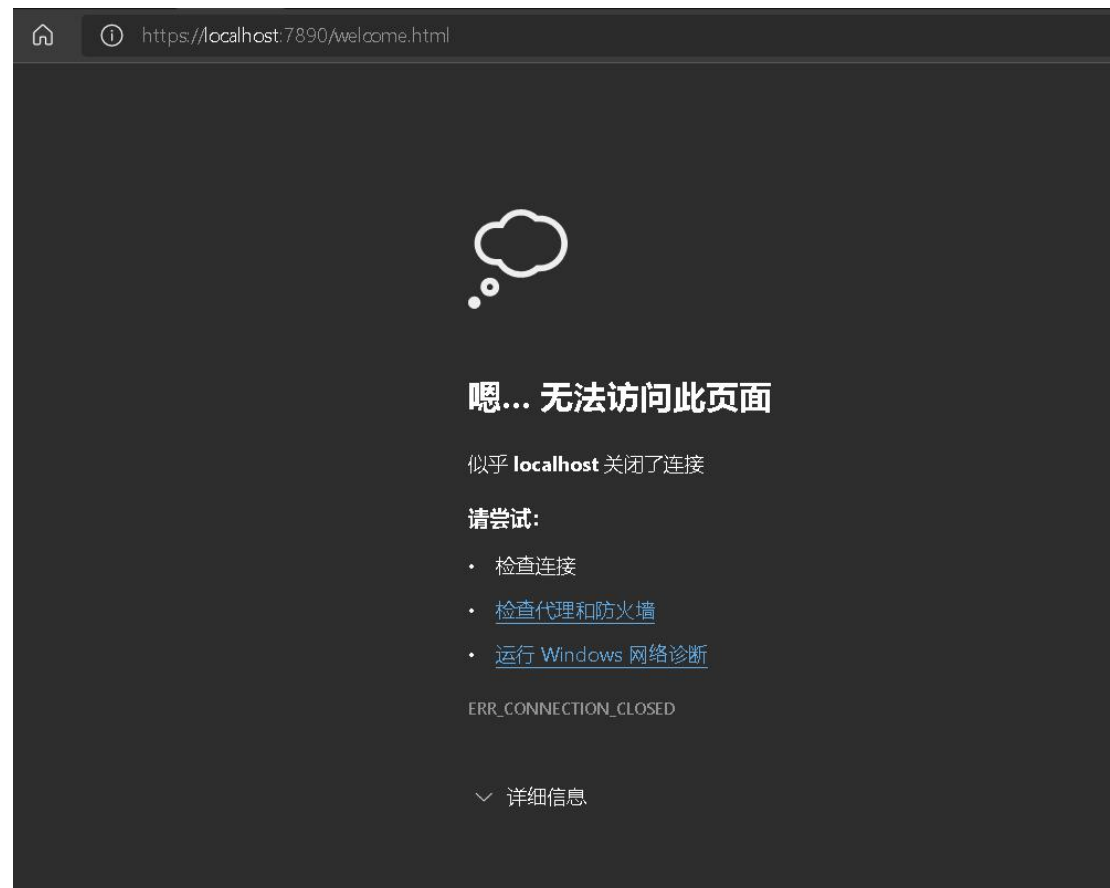
HTTPS（Hypertext Transfer Protocol Secure: 超文本传输安全协议）是一种透过计算机网络进行安全通信的传输协议。HTTPS 经由 HTTP 进行通信，但利用 SSL/TLS 来加密数据包。HTTPS 开发的主要目的，是提供对网站服务器的身份认证，保护交换数据的隐私与完整性。联系报错提示信息，可以推测，https 协议由于使用的不是明文，因此 decode("utf-8")不能解析请求。

验证:

1、将 index.html 中关于 welcome.html 的请求修改为 https 协议

```
<body>
</br>
<div style="text-align: center;"><h2>Connect to local web server!</h2></div>
<div style="text-align: center;"></div>
</br>
<div style="text-align: center;"><h4><a href="https://localhost:7890/welcome.html" target="_blank">welcome</a></h4></div>
</body>
```

2、先访问 index.html，可以正常显示，再点击 welcome 访问 welcome.html，显示无法访问



报错信息如下

```
Exception in thread Thread-9:
Traceback (most recent call last):
  File "C:\Users\admin\AppData\Local\Programs\Python\Python36\lib\threading.py", line 916, in _bootstrap_inner
    self.run()
  File "C:\Users\admin\AppData\Local\Programs\Python\Python36\lib\threading.py", line 864, in run
    self._target(*self._args, **self._kwargs)
  File "D:/资料/webserver/src/webserver4.py", line 11, in service_client
    request = new_socket.recv(1024).decode("utf-8")
UnicodeDecodeError: 'utf-8' codec can't decode byte 0xfc in position 8: invalid start byte
```

3、改为 http 协议，可正常跳转

```
</br>
<div style="text-align: center;"><h4><a href="http://localhost:7890/welcome.html" target="_blank">welcome</a></h4></div>
</body>
```

Nice to meet you!!!

验证成功