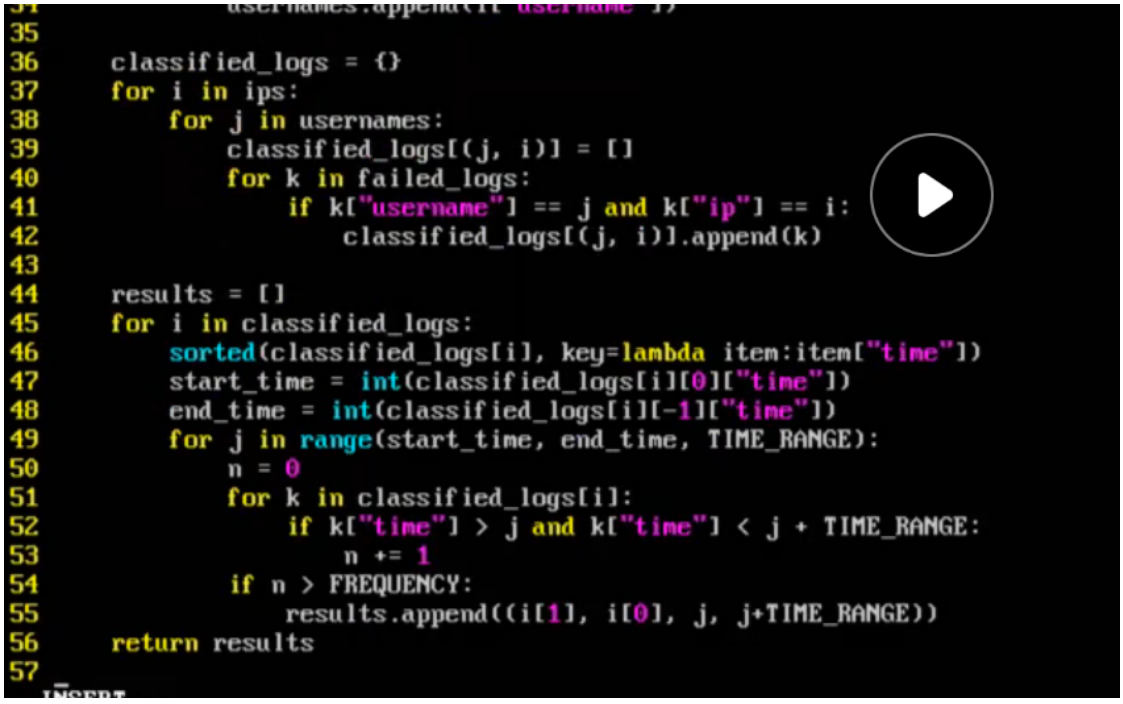
20200820信息安全实训

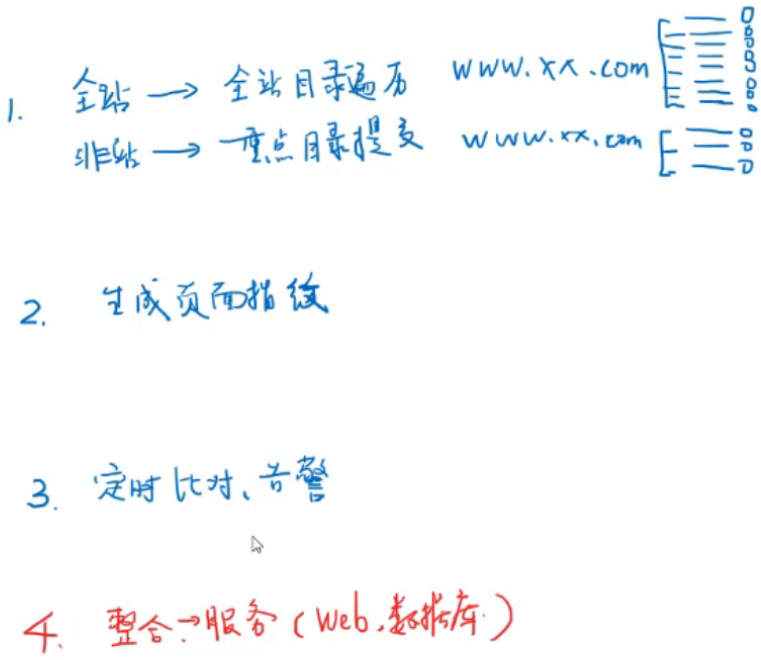
笔记本: 我的第一个笔记本
创建时间: 2020/8/20 17:36 **更新时间:** 2020/8/20 18:46
作者: 820410740@qq.com



```
35
36      classified_logs = {}
37      for i in ips:
38          for j in usernames:
39              classified_logs[(j, i)] = []
40              for k in failed_logs:
41                  if k["username"] == j and k["ip"] == i:
42                      classified_logs[(j, i)].append(k)
43
44      results = []
45      for i in classified_logs:
46          sorted(classified_logs[i], key=lambda item:item["time"])
47          start_time = int(classified_logs[i][0]["time"])
48          end_time = int(classified_logs[i][-1]["time"])
49          for j in range(start_time, end_time, TIME_RANGE):
50              n = 0
51              for k in classified_logs[i]:
52                  if k["time"] > j and k["time"] < j + TIME_RANGE:
53                      n += 1
54              if n > FREQUENCY:
55                  results.append((i[1], i[0], j, j+TIME_RANGE))
56      return results
57
```

正则表达式

作业:



可用性监测

获取全站链接

```python
import requests
import re

res = requests.get("https://www.szlanyou.com/", verify=False)

regexp = 'href="'

links = re.findall(regexp, res.text)

print(links)
```

Beautiful Soup

拼接域名
urllib

```python
from bs4 import BeautifulSoup
import requests
from urllib.parse import urljoin

URL = "https://www.szlanyou.com"

res = requests.get(URL, verify=False)

soup = BeautifulSoup(res.text, 'html.parser')

for link in soup.find_all('a'):
    href = link.get("href")
    if URL in href:
        print(href)
```

```python
from bs4 import BeautifulSoup
import requests
from urllib.parse import urljoin
from urllib.parse import urlparse

URL = "https://www.szlanyou.com"

res = requests.get(URL, verify=False)

soup = BeautifulSoup(res.text, 'html.parser')

hrefs = set()
for link in soup.find_all('a'):
    href = link.get("href")
    href_parse = urlparse(href)
    if URL in href:
        hrefs.add(href)
    elif href_parse[1] == '' and 'javascript' not in href:
        hrefs.add(urljoin(URL, href))

for i in hrefs:
    print(i)
```