第二部分

我选择的第一个问题是(请将不做的题目编号删除)： E

Pre-Task 1.1:

该部分，我选择的网站是：http://www.4399.com
Python 的具体代码如下：

```python
#导入requests库
import requests
#使用get请求
html = requests.get("http://www.4399.com/")
#判断请求是否成功
assert html.status_code == 200
#确定编码方式
html.encoding = "utf-8"
#显示结果
print(html.text)
```

最终可成功获得结果。运行结果命名为 4399.txt，存于附件。

Pre-Task 2.1:

这个正则表达式用于判断 email 格式：

\^：匹配字符串的开头
[\w._%+-]+：匹配一个或多个字符（由最后一个+表示），可以是字母、数字、下划线（\w），句点(.)、百分号(%)、加号(+)或短划线(-)。
@：匹配一个 "@" 符号
[\w.-]+：匹配一个或多个字符，可以是字母、数字、下划线、句点或短划线。
[a-zA-Z]{2,4}：匹配两到四个大小写字母。
$：匹配输入字符串的结尾位置
g：匹配字符串中所有符合条件的子串。

Pre-Task 2.2:

该部分代码实现如下:

```python
# -*- codeing = utf-8 -*-
# @Time : 2023/3/21 16:47
import re

def number(s):
    judge = r"-?\d+"  # 匹配整数或带负号的整数
    number = re.findall(judge, s)    # 使用re库找出数字
    return [int(a) for a in number]
print(number('abc123j120c0-1'))  #此处也可填入其他字符
```

显示的结果：

```
D:\python\venv\Scripts\python.exe "D:\python\Pre-Task 2.2.py"
[123, 120, 0, -1]

Process finished with exit code 0
```

Pre-Task 3.1

Bs4 中，节点之间为树形结构，主要有三种关系：

1. 父子关系：一个节点包含其它节点，这种关系为父子关系。被包含的节点称为子节点，可使用 parent 属性获取其父节点；包含其它节点的点称为父节点，可通过 children 属性获取其子节点，descendants 属性获取所有后代节点。
2. 兄弟关系：指在同一级别下的节点间的关系，它们具有相同的父节点。可使用 next_sibling 和 previous_sibling 属性获取节点的下一个兄弟节点和上一个兄弟节点。
3. 前后关系：在同一层级中，节点按照它们在 HTML 文档中出现的顺序依次排列，前一个节点是其后一个节点的前驱节点，后一个节点是其前一个节点的后继节点，这种关系为前后关系。

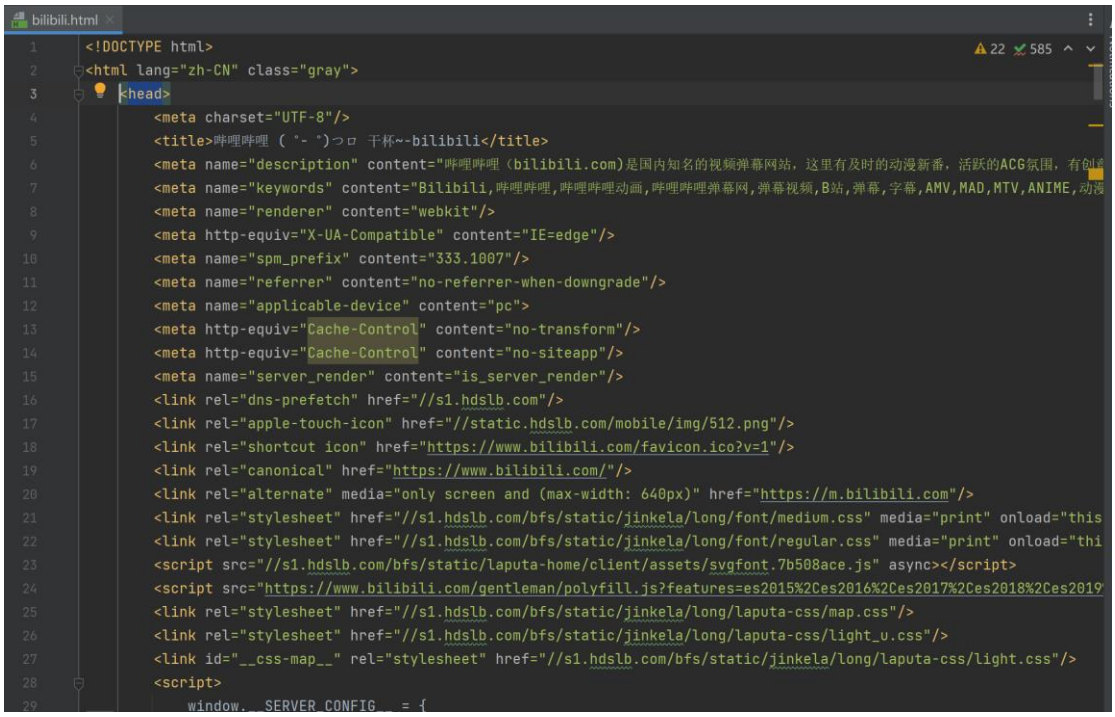
.next_element 用于获取解析过程中下一个被解析的对象。如果当前节点后面没有节点，.next_element 方法会返回 None。

Pre-Task 3.2

打开网址 https://www.bilibili.com, 单击右键

选择查看网页源代码，并复制。随后在 python 项目下创建一个 html 文件，将源代码拷贝。

结果如图：



随后编写 python 程序查找链接(编写的程序与 bilibili.html 文件应放置于同一项目)。

代码如下：

```python
# -*- codeing = utf-8 -*-
# @Time : 2023/3/20 19:34
import re
from bs4 import BeautifulSoup    #导入需要的库

file = open("./bilibili.html","rb")    #打开html文件
html = file.read()
soup = BeautifulSoup(html, 'html.parser')
result = soup.find_all("a",href=re.compile(r'^https://www\.bilibili\.com/video/BV[\w\d]+'))#运用find_all方法查找
for item in result:
    print(item['href'])        #输出
```

参考资料：

[Python 课程天花板,Python 入门+Python 爬虫+Python 数据分析 5 天项目实操/Python 基础.Python 教程 哔哩哔哩 bilibili]

[bs4 的简单介绍_举个丽子<>!!的博客-CSDN 博客]

[python——正则表达式(re 模块)详解_python re 正则表达式_nee~的博客-CSDN 博客]