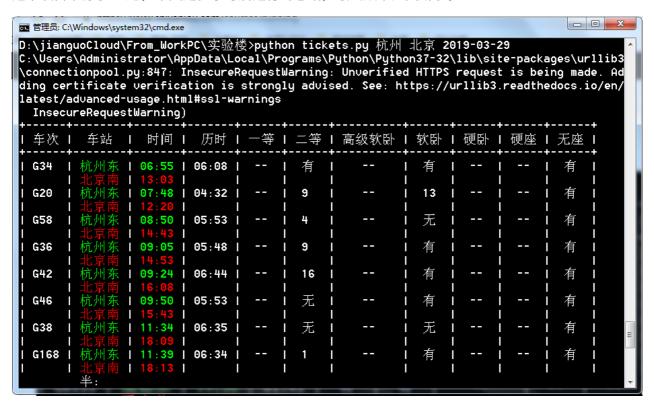
# 一、引言

学了 Python 基础语法后想锻炼一下解决实际问题的水平,那本文介绍的小项目刚刚好。通过本文中项目,你可以学会不打开12306或者APP就能查询火车票,同时熟悉 Python 库的使用方法,了解 Python 程序设计思路。

这个项目来源于 此处,本文是在学习后进行的总结,最终效果如图所示。



本项目需要 Python 基础语法知识, 在项目中可以学习到如下内容:

- docopt 、 requests 、 colorama 和 prettytable 库的使用
- 命令行安装需要的 Python 库
- 分析网站的请求和响应过程

## 二、设计思路

为了实现如图所示功能效果,我们需要先进行需求分析。

- 当我们查询火车票时一般来说需要指定条件,例如列车时间、起始站点等信息才能查看列车时刻信息,因此,我们需要明确定义有哪些查询条件;
- 然后需要有命令行界面、用来接受用户输入的查询条件;
- 根据用户输入的查询条件,去寻找列车时刻信息;
- 最后将获得的列车时刻信息经过排版再显示出来。

这样, 总结下来可以分为如下几个大的模块

- 命令行界面,用于接受用户输入的查询条件
- 列车信息API, 向12306服务器发起查询请求, 并获得列车信息
- 列车信息展示,美观的展示出查询到达列车信息

# 三、知识点

### 1.项目所需 Python 库

- docopt 库:终端命令行参数解析器,查看详情;
- requests 库: 简单易用的Python HTTP库,快速上手文档;
- prettytable 库:格式化输出打印内容,查看详情;
- colorama 库: 命令行着色工具, 查看详情;

#### 2.命令行一键安装

通过如下命令一键安装以上资源:

pip3 install requests prettytable docopt colorama

# 四、分步讲解

下面开始构建属于自己的命令行列车时刻查询工具。

#### 1.命令行界面

```
# coding: utf-8
"""命令行火车票查看器
Usage:
   tickets [-gdtkz] <from> <to> <date>
Options:
   -h,--help 显示帮助菜单
           高铁
动车
   -g
   -d
   -t
            特快
   -k
           快速
            直达
   -Z
Example:
  tickets 杭州 北京 2019-03-25
   tickets -dg 杭州 北京 2019-03-25
from docopt import docopt
def cli():
```

```
"""command-line interface"""
arguments = docopt(__doc__)
print(arguments)

if __name__ == '__main__':
    cli()
```

根据官方文档说明, docopt 需要1个必传参数和4个可选参数:

```
docopt(doc, argv=None, help=True, version=None, options_first=False)
```

- doc 是由字符串或者注释构成的命令行模块,格式如以上代码所示( doc could be a module docstring (doc) or some other string that contains a help message that will be parsed to create the option parser)。
- 其返回值类型为字典,也就是 arguments ,键为 Usage 和 Options 中选项,值为实际输入内容。
- 实际运行下如上代码,看到如下所示返回结果内容。

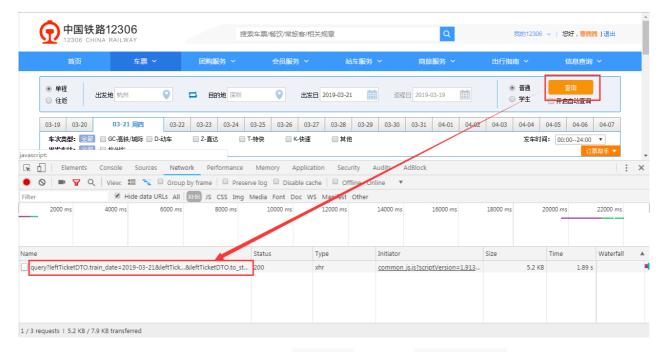
```
■ 管理员: 命令提示符
                       文件 7,338 寸 //
目录 17,980,772,352 可用字节
                                        7,330 字节
C:∖>python tickets.py
Usage:
     tickets [-gdtkz] <from> <to> <date>
C:↘python tickets.py -h
命令行火车票查看器
Usage:
    tickets [-gdtkz] (from) (to) (date)
Options:
                   显示帮助菜单
高铁
动车
特快
    -h,--help
     -\mathbf{g}
     -\mathbf{d}
     -t
     -\mathbf{k}
    -z
Example:
    tickets 杭州 北京 2019-03-25
tickets -dg 杭州 北京 2019-03-25
C:<>python tickets.py -gd 北京 杭州 2019-03-25
<'-d': True,
 '-g': True,
 '-k': False,
 '-t': False,
 '-z': False,
 '<date>': '2019-03-25',
'<from>': '北京',
'<to>': '杭州'>
C: 🖴
4
                                             -111
```

#### 2.获取列车信息

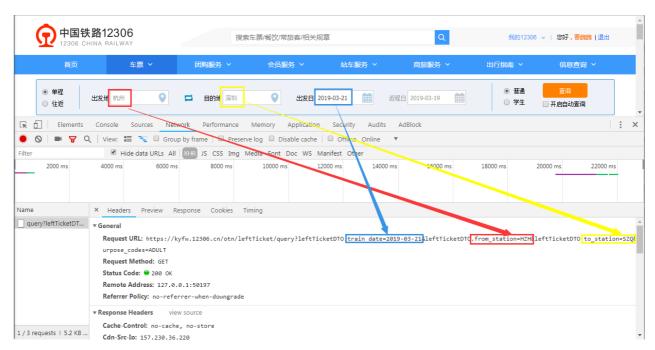
查火车票只能去12306,但是铁道部没有提供API,那么我们就不能查询了吗,答案是否定的。我们现在 在网站上查询几次列车信息,寻找接口的蛛丝马迹。

#### 2.1分析页面接口

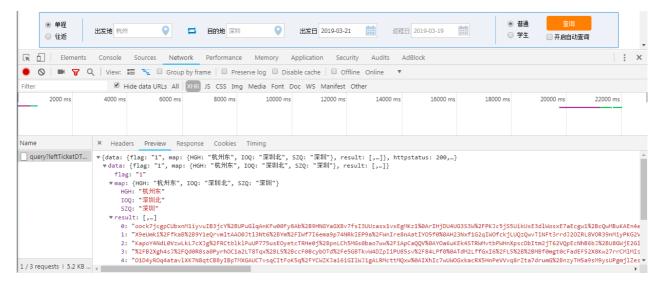
• 在列车查询页面,然后打开浏览器调试模式(Windows下Chrome浏览器快捷键是F12),进入 Network 中,查看 XHR ,然后查询列车信息



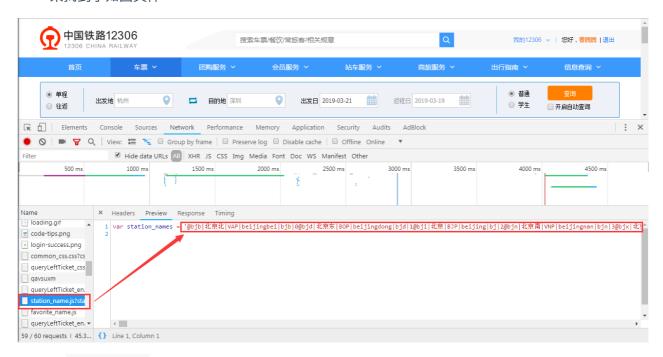
• 点击箭头指向的文件,我们可以看到在 Headers 栏目中的 Request URL 中包括了查询的所有条件信息: 出发地、目的地、出发日



• 再点击 Preview 查看返回结果,这个就是查询到的列车时刻信息了



知道如何查询列车车次信息后,还有一个问题,就是我们输入的是中文车站名,接口处理的是英文车站代码,这个映射关系还不知道,需要继续查看页面加载的文件,寻找这个对应关系的文件,结果找到了如图文件



#### 2.2使用 requests 库

接口信息分析完毕,下面用 requests 库进行实战。

• 首先,来获得我们想要的车站英文代码数据,通过向车站中英文关系文件发起请求获得响应,分析 结果

```
# --coding: UTF-8--
import re, requests
from pprint import pprint

# 车站中英文文件URL
url = 'https://kyfw.12306.cn/otn/resources/js/framework/station_name.js?
station_version=1.9098'

# 使用requests库发起请求, 获得响应的json文件
resp = requests.get(url)
```

```
# 对获得的响应文件使用正则表达式,获得车站中文名和英文代码组成的元祖构成的列表 stations = re.findall(r'([\u4e00-\u9fa5]+)\|([A-Z]+)', resp.text) # 格式化输出车站 pprint(dict(stations), indent=4)
```

- stations 返回的结果如下图,是车站中文名和英文代码构成的元组组成的列表



— 使用 python parse\_stations.py > stations.py 命令并生成 stations.py 文件, 车站文件 内容为中文名和英文代码构成的字典, 这样就可以导入 stations.py 文件后根据键(车站中文名)从 字典中检索对应值(车站中文对应的英文代码)

- 回顾一下,在命令行界面输入的起始车站中文名和列车出发日期,这些信息都在 arguments 这个字典对象中,只需要根据输入的信息匹配车站英文代码,然后访问车次查询接口,这样就可以查询 列车信息了
  - 修改命令行函数 cli 中代码

```
def cli():
    """command-line interface"""
    arguments = docopt(__doc__)
    # 根据命令行用输入的起始车站名称、日期,查询stations中对应的英文代码
    from_station = stations.get(arguments['<from>'])
    to_station = stations.get(arguments['<to>'])
    date = arguments['<date>']

# 发起请求查询车次信息
    url = 'https://kyfw.12306.cn/otn/leftTicket/query?leftTicketDTO.train_date=
{}&leftTicketDTO.from_station={}&leftTicketDTO.to_station=
{}&purpose_codes=ADULT'.format(date, from_station, to_station)
    r = requests.get(url)
    # 将返回的json数据转为字典类型
    # 其中车次信息在返回的字典键为data的值嵌套的字典中,其嵌套的键为result的值中
```

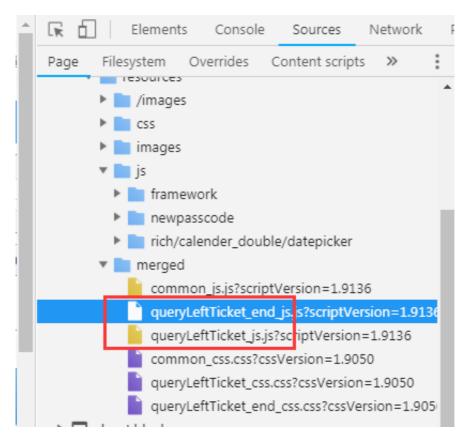
```
available_trains = r.json()['data']['result']
```

#### 返回的车次结果信息如图所示

#### 2.3解析车次信息

对获取的车次信息一头雾水,那么需要尝试继续在网站中寻找线索。一般来说,网站的js文件中会定义 这些数据结构,并标注每个字段的含义,所以现在来寻找这样的js文件。

进入浏览器调试模式,查看源代码 source 。猜测是图中所示的文件可能包括车票信息,因为查询车次就是查询列车余票信息。



将压缩的js代码调整下格式,然后尝试搜索"硬座"与列车相关的信息关键词,结果找到了如下代码,这下知道坐席代码了:

```
var yxTrainPageSize = 15;
var passengerPageSize = 20;
var timer_time = 3;
var yxTrainChange = "";
var trainListForIE = [];
var queryLeftTicket_times = 0;
var queryLeftTicket_count = 10;
var ifAlertCode = false;
var intervalTime;
var seatlypeForHB = {
    SWZ: "9_商务座",
    TZ: "P_特等座",
    ZY: "M_一等座",
    ZY: "M_一等座",
    ZY: "M_一等座",
    ZR: "0_二等座",
    RW: "4_软卧",
    SRRB: "F_动卧",
    YW: "3_硬卧",
    YW: "3_硬卧",
    YZ: "1_硬座",
    VZ: "1_天座",
    QT: "H_其他"
    };
var _alowHBHour = 19;
var alowHBMaxNum = 2;
```

继续寻找"硬座",发现并没有更多有价值内容,然后尝试搜索"硬座"的键值"YZ",发现关键代码:

```
function cp(c0, cQ) {
      var cN = [];
for (var cM = 0; cM < c0.length; cM++) {
              var cR = [];
              var cL = c0[cM].split("|");
              cR.secretStr
                          cR.buttonTextInfo = cL[1];
                          var cP = [];
                          cP.train_no = cL[2];
                          cP.station_train_code = cL[3];
                          cP.start_station_telecode = cL[4];
                          cP.end_station_telecode = cL[5];
                          cP.from_station_telecode = cL[6];
                          cP.to_station_telecode = cL[7];
                          cP.start_time = cL[8
                          cP.arrive_time = cL[9];
cP.lishi = cL[10];
                          cP.canWebBuy = cL[11];
                          cP.yp_info = cL[12];
                          cP.start_train_date = cL[13];
                          cP.train_seat_feature = cL[14];
                          cP.location_code = cL[15];
                          cP.from_station_no = cL[16];
                          cP.to_station_no = cL[17];
cP.is_support_card = cL[18];
                          cP.to_station_no = cL[17
                         cP.is_support_card = cL[18];
cP.controlled_train_flag = cL[19];
cP.gg_num = cL[20] ? cL[20] : "--";
cP.gr_num = cL[21] ? cL[21] : "--";
cP.qt_num = cL[22] ? cL[22] : "--";
cP.rw_num = cL[23] ? cL[23] : "--";
cP.rz_num = cL[24] ? cL[24] : "--";
cP.wz_num = cL[25] ? cL[25] : "--";
cP.wz_num = cL[26] ? cL[26] : "--";
cP.ww_num = cL[28] ? cL[28] : "--"
                                                                               "--"
                         CP.yb_num = cL[27] ? cL[27] : "--";
cP.yw_num = cL[28] ? cL[28] : "--";
cP.= _num = cL[29] ? cL[29] : "--";
cP.ze_num = cL[30] ? cL[30] : "--";
cP.zy_num = cL[31] ? cL[31] : "--";
cP.swz_num = cL[32] ? cL[32] : "--";
cP.srrb_num = cL[33] ? cL[33] : "--";
cP.yp_ex = cL[34];
cP.seat_types = cL[35]:
                          cP.seat_types = cL[35];
                          cP.exchange_train_flag = cL[36];
                          cP.houbu_train_flag = cL[37];
                          if (cL.length > 38) {
                                 cP.houbu_seat_limit = cL[38]
                          cP.from_station_name = cQ[cL[6]];
                          cP.to_station_name = cQ[cL[7]];
cR.queryLeftNewDTO = cP;
                          cN. push(cR)
                   return cN
```

完美,这段代码十分详细的介绍了车次信息中各个值的含义,我们再拿来和在12306查询到的车次信息 比对发现完全正确,这样就可以解析车次信息了。

## 3.列车信息展示

#### 3.1构建列车信息类

接下来使用类来封装获取的列车信息,整体思路是,通过类来处理页面爬虫获取的列车信息,并将数据清洗后展示出来,列车信息类 TrainsInfo 的代码如下:

```
class TrainsInfo:
    def __init__(self, available_trains, from_to_stations, options):
        """
        定义类用到的变量: 列车信息,列车类型选项
        available_trains: 列车信息
```

```
fom_o_saions: 起止车站,模糊匹配,例如杭州,杭州东等
       options: 列车类型选项
       0.0000
       self.available_trains = available_trains
       self.from_to_stations = from_to_stations
       self.options = options
   @property
   def trains(self):
       for each_train in self.available_trains:
           each_train_split = each_train.split('|')
           train_code = each_train_split[3]#车次
           if not self.options or train_code[0].lower() in self.options):
               显示符合条件的列车信息,即列车类型匹配g/d/z等 或者
               不输入列车等级 也可查询
               trains = [train_code,
                                          #车次
                   each_train_split[6],
                                         #起点车站代码
                   each_train_split[7],
                                         #终点车站代码
                   each_train_split[8],
                                         #发车时间
                   each_train_split[9],
                                         #到达时间
                   each_train_split[10],
                                         #历时
                   each_train_split[11], #网上购票
                   each_train_split[26] if each_train_split[26] else '--',
                                                                            #无座
                   each_train_split[27] if each_train_split[27] else '--',
                                                                            #硬座
                   each_train_split[24] if each_train_split[24] else '--',
                                                                            #软座
                   each_train_split[28] if each_train_split[28] else '--',
                                                                            #硬卧
                   each_train_split[33] if each_train_split[33] else '--',
                                                                            #动卧
                   each_train_split[23] if each_train_split[23] else '--',
                                                                            #软卧
                   each_train_split[21] if each_train_split[21] else '--',
                                                                            #高级软
卧
                   each_train_split[30] if each_train_split[30] else '--',
                                                                            #二等座
                   each_train_split[31] if each_train_split[31] else '--',
                                                                            #一等座
                   each_train_split[25] if each_train_split[25] else '--',
                                                                            #特等座
                   each_train_split[32] if each_train_split[32] else '--',
                                                                            #商务座
                   each_train_split[22] if each_train_split[22] else '--',
                                                                            #其他
               yield trains
   def print_train_info(self):
       for train in self.trains:
           print(train)
```

为了调用类,修改后的 cli 代码如下:

```
def cli():
    """command-line interface"""
```

```
arguments = docopt(__doc__)
   from_station = stations.get(arguments['<from>'])
   to_station = stations.get(arguments['<to>'])
   date = arguments['<date>']
   """获取查询得到的列车车次及起止车站信息"""
   url = 'https://kyfw.12306.cn/otn/leftTicket/query?leftTicketDTO.train_date=
{}&leftTicketDTO.from_station={}&leftTicketDTO.to_station=
{}&purpose_codes=ADULT'.format(date, from_station, to_station)
   r = requests.get(url)
   available_trains 获取查询到的列车信息,返回值是车次信息列表
   from_to_stations 获取起止车站信息,返回值是车站英文代码和车站中文构成的字典
   options 获取从命令行输入的参数,将列表转为字符串,方便后续使用
   available_trains = r.json()['data']['result']
   from_to_stations = r.json()['data']['map']
   options = ''.join([key for key, value in arguments.items() if value != False])
   TrainsInfo(available\_trains, from\_to\_stations, options).print\_train\_info()
```

测试下输出的结果,输入 python tickets.py 杭州 绩溪县 2019-03-30 ,可查询到如下列车信息:

```
■ 管理员: C:\Windows\system32\cmd.exe
C:\>python tickets.py 杭州 绩溪县 2019-03-30
                         08:00',
                                  '09:33', '01:33', 'Y', '--'
                         . '11',
'08:07'
                                  '09:53', '01:46', 'Y', '--
                                   24:00', '99:59', 'IS_TIME_NOT_BUY',
  G2379'
                         24:00
                                  14:24', '01:32',
                                                     'Y', '3', '--',
 G7191
          'HGH'
                  'NRH
                         12:52
                         2'.
                                   16:50', '01:39', 'Y', '有', '--', '--',
 G7315
                                  '17:35', '01:26', 'Y', '有', '--',
                          16:09
 D3397
                  NRH
                          16:19
                                  '17:52', '01:33', 'Y', '无', '--',
           HGH '
                          16:47
                                  '18:15', '01:28', 'Y', '--', '--', '--',
 G2381
 G7303'
                          17:34
                                  '19:06', '01:32', 'Y', '20', '--', '--', '--',
                           5.
                                  '19:43', '01:33', 'Y', '有', '--', '--',
                          18:10
           HGH'
                                  '20:13', '01:20', 'Y', '--', '--',
 G1393
                          18:53
                                  '20:39', '01:26', 'Y', '有', '--'
 G7307
                          19:13
                                  '20:51', '01:27', 'Y', '有', '--',
                          19:24
                                  '21:42', '01:32', 'Y', '有', '--',
C:\>
```

看一下12306的查询结果,比对一下,完全一致。

车次	出发站 到达站	出发时间▲ 到达时间▼	历时 🐣	商务座 特等座	一等座	二等座	高級 软卧	牧卧 一等卧	动卧	個斯 個等二	软座	硬座	无座	其他	备注
<u>G2385</u> •	□ 杭州东 □ 练奚北	08:00 09:33	<b>01:33</b> 当日到达	11	有	有		-	-	-	-		-	-	预订
<u>G2395</u> 🗸	№ 杭州东 ७ 绩奚北	08:07 09:53	<b>01:46</b> 当日到达	无	无	2		_	-	-	-			_	预订
<u>G7191</u> •	□ 杭州东 □ 绩奚北	12:52 14:24	<b>01:32</b> 当日到达	2	1	无		-	-	-			15	-	预订
<u>G7315</u> •	図 杭州东 図 绩奚北	15:11 16:50	<b>01:39</b> 当日到达	18	有	有		_	_	-			有	_	预订
<u>D2109</u> •	№ 杭州东 閏 绩奚北	16:09 17:35	<b>01:26</b> 当日到达	-	19	有		-	-	-	-		有	-	预订
<u>D3397</u>	№ 杭州东 ⑩ 绩奚北	16:19 17:52	<b>01:33</b> 当日到达	-	有	有		_	-	-			无	_	预订
<u>G2381</u> •	№ 杭州东 ⑩ 绩奚北	16:47 18:15	<b>01:28</b> 当日到达	无	无	有		-	-	-				-	预订
<u>G7303</u> •	図 杭州东 図 绩奚北	17:34 19:06	<b>01:32</b> 当日到达	5	19	有		_	-	-			20	_	预订
<u>G7451</u> •	□ 杭州东 □ 绩奚北	18:10 19:43	<b>01:33</b> 当日到达	8	有	有		-	-	-			有	-	预订
<u>G1393</u> •	図 杭州东 図 绩奚北	18:53 20:13	<b>01:20</b> 当日到达	无	无	有		-	-	-				-	预订
<u>G7307</u> •	□ 杭州东 □ 绩奚北	19:13 20:39	<b>01:26</b> 当日到达	3	19	有		-	-	-	-		有	-	预订
<u>D5585</u> •	№ 杭州东 図 绩奚北	19:24 20:51	<b>01:27</b> 当日到达	5	20	有		_	_	-	-		有	_	预订
D3323 •	□ 杭州东 □ 绩奚北	20:10 21:42	<b>01:32</b> 当日到达	-	有	有	-	_	_	_	_	-	有	_	预订
<u>G2379</u>	22 杭州东 22 绩奚北			-	_			_	_	_	-		_	_	列车停运

#### 3.2美化输出界面

但是工作到这里并没有结束,这个输出界面太丑陋了,需要精致一些。使用prettytable库来将信息用表格形式输出,在调整些颜色即可。

调整列车信息打印函数,通过表格输出信息,在使用前需要引用 prettytable 库 import prettytable

```
def print_train_info(self):
    # 创建表格
    tb = prettytable.PrettyTable()
    # 设置表格标题
    tb.field_names = self.headers
    for train in self.trains:
        # 增加每行数据
        tb.add_row(train)
    print(tb)
```

列车信息类中需要对train信息进行颜色标识,整体代码修改后如下:

```
init()

class TrainsInfo:
    headers = '车次 车站 时间 历时 有票 无座 硬座 软座 硬卧 动卧 软卧 高级软卧 二等座 一等座 特等 座 商务座 其他'.split()

def __init__(self, available_trains, from_to_stations, options):
    """
    定义类用到的变量: 列车信息, 列车类型选项
    available_trains: 列车信息
    fom_o_saions: 起止车站, 模糊匹配, 例如杭州, 杭州东等
    options: 列车类型选项
    """
```

```
self.available_trains = available_trains
       self.from_to_stations = from_to_stations
       self.options = options
   @property
   def trains(self):
       for each_train in self.available_trains:
           each_train_split = each_train.split('|')
           train_code = each_train_split[3]#车次
           if not self.options or train_code[0].lower() in self.options:
               显示符合条件的列车信息,即列车类型匹配g/d/z等 或者
               不输入列车等级 也可查询
               trains = [train_code,
                                          #车次
                   '\n'.join([Fore.GREEN + self.from_to_stations[each_train_split[6]]
+ Fore.RESET,
                              Fore.RED + self.from_to_stations[each_train_split[7]] +
Fore.RESET]),
                   '\n'.join([Fore.GREEN + each_train_split[8] + Fore.RESET,
                              Fore.RED + each_train_split[9] + Fore.RESET]),
                   each_train_split[10], #历时
                   each_train_split[11], #网上购票
                   each_train_split[26] if each_train_split[26] else '--',
                                                                              #无座
                   each_train_split[27] if each_train_split[27] else '--',
                                                                              #硬座
                   each_train_split[24] if each_train_split[24] else '--',
                                                                              #软座
                   each_train_split[28] if each_train_split[28] else '--',
                                                                              #硬卧
                   each_train_split[33] if each_train_split[33] else '--',
                                                                              #动卧
                   each_train_split[23] if each_train_split[23] else '--',
                                                                              #软卧
                   each_train_split[21] if each_train_split[21] else '--',
                                                                              #高级软
卧
                   each_train_split[30] if each_train_split[30] else '--',
                                                                              #二等座
                   each_train_split[31] if each_train_split[31] else '--',
                                                                              #一等座
                   each_train_split[25] if each_train_split[25] else '--',
                                                                              #特等座
                   each_train_split[32] if each_train_split[32] else '--',
                                                                              #商务座
                   each_train_split[22] if each_train_split[22] else '--',
                                                                              #其他
               yield trains
```

这里面用到了 colorama 命令行着色工具,使用前需要引入库 from colorama import init, Fore, 然后初始化 init() 并设置颜色 Fore.GREEN Fore.RED Fore.RESET

## 五、总结

本项目最核心知识点在于 requests 库的使用,难点在于如何分析js文件寻找车站和英文关系对应以及 列车车次信息获取后如何解析。 其他的知识点还有 yield 使用, @property 装饰器, dicopt 命令行界面以及 prettytable 表格输出 colorama 表格上色。

详细代码请查阅: https://github.com/elicao/12306query-tickets