机票比价程序

选题立意

> 网络的迅猛发展将人类带入了信息社会和网络经济时代，这对个人生活产生了深刻的影响。尤其在出行方面，用户可以通过很多不同的网站在线上就能选择自己适宜的出行方式，但如果用户想要选择最具有性价比的出行方式，就需要自身花费大量时间来浏览大量网站。因此用户需要一个能够比较各网站出行信息功能的程序，基于现实需要以及对自身技能的锤炼和检验，故选择出行中的一方面飞机机票的比价程序作为网络爬虫课程结课作业的题目。

目标设定

1. 选择爬取网站及内容

> 1. 选择携程、飞猪以及去哪儿网的机票页面作为爬取内容。

2. 基本功能

> 1. 根据航班起始地、目的地、航班日期作为条件,爬取三家网站中符合条件的航班。
> 2. 根据机票价格对航班进行排序，从而得知当天符合条件航班中最大最小机票价格的航班。
> 3. 对同一航班号的不同网站的机票价格进行对比。

3. 补充功能

> 1. 实现网站的登录。
> 2. 对当前符合条件的航班进行各种排序及分类。
> 3. 实现根据航班号进行订票操作。

具体方法实现

1. 选定基本技术路线

> 1. 通过分析携程、飞猪以及去哪网的机票查询页面得知这些页面都是动态页面，所以使用selenium作为主要技术手段与网页交互。
> 2. 但selenium获取网页中的元素信息较慢,使用beautifulsoup作为辅助技术手段用于获取网页信息。

2. 如何传递查询参数给网站,使之查询相应区间的机票

> 1. 飞猪、去哪儿网可直接给链接传递时间、起始地、目的地等参数，可直接跳转到相应航班机票的查询页面
>
>
> ```
> fliggyUrl=f"https://sjipiao.fliggy.com/flight_search_result.htm?_input_charset=utf-8&depCityName={fromCity}&depDate={fromDate}&arrCityName={reachCity}"
> qunarUrl=f"https://flight.qunar.com/site/oneway_list.htm?searchDepartureAirport={fromCity}&searchArrivalAirport={reachCity}&searchDepartureTime={fromDate}"
> ```

2. 携程也可以通过链接传递相应参数,跳转到相应航班的机票页面,但携程的起始地和目的地的参数都需要使用地点的对应id指定.在本项目中,是通过与页面交互来获得地点id的。

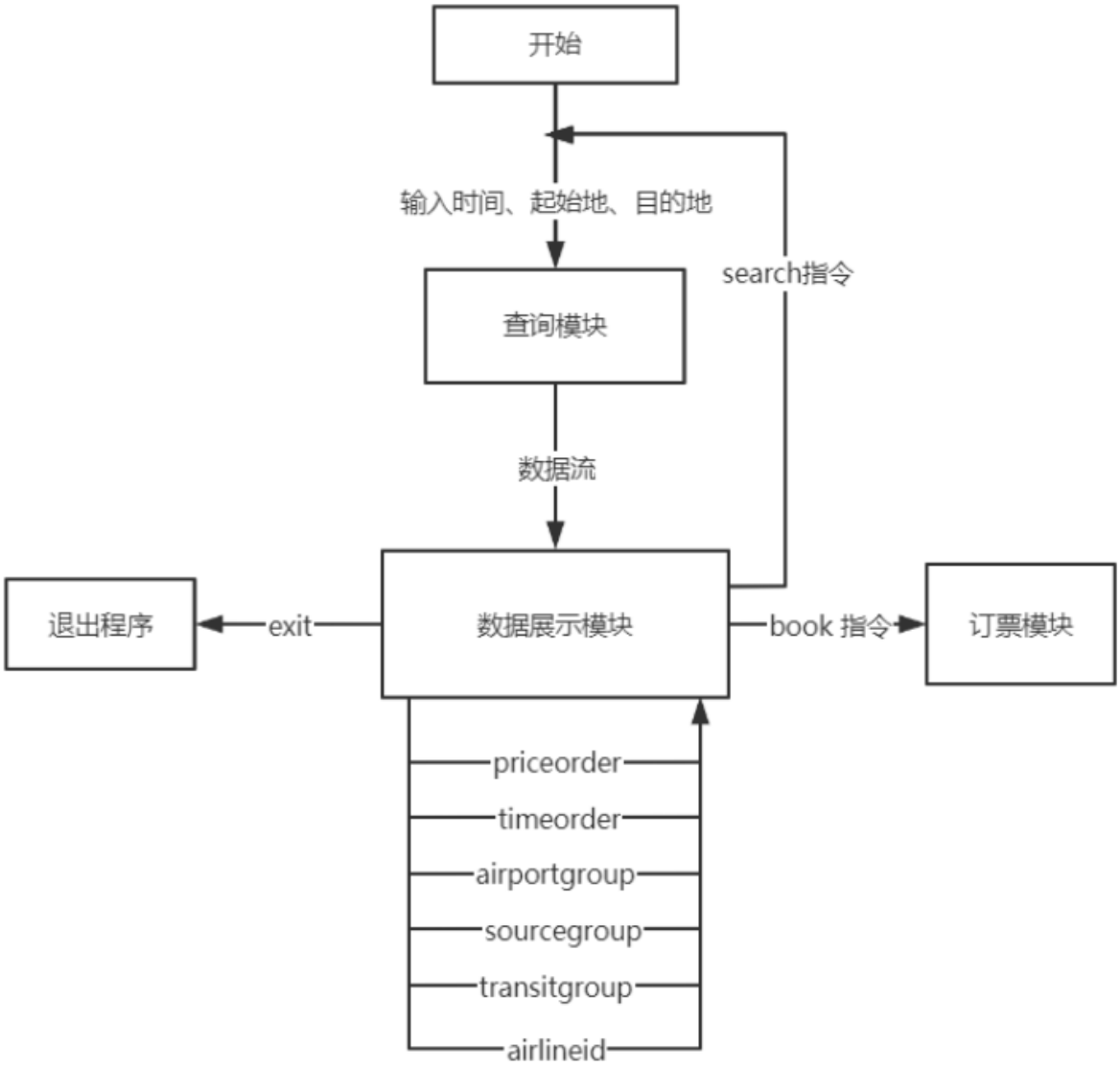
```
fromCityE.send_keys(fromCity)
reachCityE.send_keys(reachCity)
fromCityId=re.search(r"[A-Z]+",fromCityE.get_attribute("value")).group().lower()
reachCityId=re.search(r"[A-Z]+",reachCityE.get_attribute("value")).group().lower()
ctripUrl=f"https://flights.ctrip.com/itinerary/oneway/{fromCityId}-{reachCityId}?date=
{fromDate}&hasChild=false&hasBaby=false&classType=ALL"
```

3. 如何从网页中获取该网站的相应航班机票信息

1. 基本流程 :用selenium WebDriverWait等待页面中某个元素加载出现然后通过driver.page_source提取出源代码给beautifulsoup进行页面中信息的获取。
2. 飞猪网使用基本流程即可获得想要的航班信息。
3. 携程网的信息分段动态加载的,需要将滚动条滑到底部才会加载全部信息,所以携程网在提取网页源代码之前需要使用selenium driver执行js嵌入将滚动条滑到底部之后,再用基本流程获取网页中的信息。
4. 去哪儿网的信息是分页加载的,需要使用seleniumdriver跳转页面,其每一页使用基本流程获取网页信息。

4. 信息整理与展示

1. 使用一个通用的airlineInfo类来进行航班信息的整理与记录,程序在网页获取信息时,每一个航班会创建一个airlineInfo的实例对象追加到列表里。
2. 将各模块在最外层封装为一个直观易用的流程。
3. 在数据展示模块中,用户通过指令与程序进行交互,实现重新搜索(search)、同航班比价(airlineid航班号)、订票（book订票id）、根据价格升降序排序(priceorder/Rorder)、根据起飞时间升降序排序(timeorder/Rorder)、根据起飞机场分类(airportgroup)等等功能的使用。
4. 同航班比价功能的实现是通过指令中携带的航班号对航班信息的列表进行遍历,获得该航班号对应的不同来源的机票的信息。
5. 根据xx升降序功能的实现就是简单的列表sort排序功能的使用,通过传递一个lambda表达式进行不同元素的排序。
6. 分类功能的实现其实也是使用列表sort排序功能的使用,对于同一个分类,使其lambda表达式得出同一个值,便可分类在一起。



5. 登录功能的实现

1. 原因: 飞猪和去哪儿网非登录的情况下只有一段时间可以查询机票信息,之后需要登录才能查询,还有就是为后面的订票功能做铺垫。
2. 去哪儿网和携程登录: 此处通过记录登录的cookie信息实现登录(因为都有验证码),但selenium进行爬取的chrome是无头(headless)隐藏不可见的(用户也不应该看到),此时通过弹出一个新的chrome浏览器,让用户进行登录,登录完成之后读取此chrome的cookie,然后将cookie传递给后台运行的headless chrome实现登录,而且程序会将该cookie记录下来,在此后登录时,便可直接通过cookie登录。
3. 飞猪网登录:其实就是淘宝登录,淘宝如果是常用账号,直接使用selenium在input中输入账号密码即可登录(无验证码),但要实现cookie信息的保存和传递,需要保存和传递淘宝和飞猪网两个网站的cookie,因为在飞猪网验证cookie时,会交叉验证当前浏览器下淘宝网的cookie信息. 。

参考资料1: 单点登录，第三方cookie和cna

6. 订票功能的实现

1. 用户使用订票功能需要传递一个bookTicketId给程序,程序根据bookTicketId可以定位到当初产生该票信息的查票网页,然后使用selenium点击相应的订票按钮,即可得到订票窗口。
2. 但爬虫的网页的selenium是以headless(无头模式)隐藏的,用户是不可见的,所以需要保存订票窗口的cookie信息,selenium打开一个新的chrome浏览器,将cookie传递给该新chrome,并使用该chrome打开

> 订票链接,用户就作这个新产生的chrome进行订票信息的补充,且不会影响后台的爬虫chrome运行。

7.网站的反爬虫策略以及如何应对

1. 飞猪网和去哪儿网在非登录的情况下只有一段时间可以查询机票信息,之后需要登录才能查询,可以通过实现cookie登录来应对。
2. 去哪儿网的价格是分位显示的,且有重叠元素是会覆盖干扰的.但可以通过识别重叠中最上层的元素从中整理出真实的价格。
3. 飞猪网对selenium有着非常严格的检测策略,如果检测到使用的是selenium登录网页,会一直弹出验证滑块,其中的一种策略是通过识别当前网页js域是否存在webdriver对象,另一种策略是通过识别webdriverjs的标识字符串 $cdc_asdjflasutopfhvcZLmcfl_。

   第一种策略的应对手段

   ```
   # 创建driver的传入相应的参数,可以达到隐藏 webdriver对象的结果
   option.add_experimental_option('excludeSwitches', ['enable-automation'])
   driver = Chrome(options=option)
   driver.execute_cdp_cmd("Page.addScriptToEvaluateOnNewDocument",{
          "source": """Object.defineProperty(navigator, 'webdriver', {get: () =>
   undefined})""",
   })
   ```

   第二种策略的应对方式

   ```
   使用Notepad++或其他代码编辑软件,打开chromedriver.exe 搜索 $cdc_asdjflasutopfhvcZLmcfl
   替换成$zyf_asdjflasutopfhvcZLmcfl
   ```

   参考资料2: 淘宝反扒解决方案

代码解析

1. airlineInfo类:航班信息的通用数据格式化类

> def **str**(self): 对数据进行整行格式化输出
>
> def tableHead(cls): 对数据表头整行格式化输出

```
import re


class airlineInfo :

    def __init__(self,airlineName,airlineId,airlineModel,flightTime,
                 reachTime,flightPort,reachPort,flightOntimeRate,
                 price,websiteName,websiteId,bookTicketId,

isTransit=False,transitLineId=None,transitAirName=None,transitAirModel=None):

        self.airlineName =airlineName
```

```
        self.airlineId =airlineId
        self.airlineModel =airlineModel
        self.flightTime =flightTime
        self.reachTime =reachTime
        self.flightPort =flightPort
        self.reachPort =reachPort
        self.flightOntimeRate =flightOntimeRate
        self.price =price
        self.websiteName=websiteName
        self.websiteId=websiteId
        self.bookTicketId=bookTicketId
        self.isTransit=isTransit
        self.transitAirName=transitAirName
        self.transitLineId=transitLineId
        self.transitAirModel=transitAirModel
    def __str__(self):
         return "%s%s%s%s%s%s%s%s%s" \
                % (containGBKStrFormat(str(self.websiteId) + str(self.airlineId),10),
                   containGBKStrFormat(self.websiteName,7),
                   containGBKStrFormat(self.airlineName if self.airlineId in
self.airlineName else self.airlineName+self.airlineId,16),
                   containGBKStrFormat(self.airlineModel,20),
                  containGBKStrFormat(self.flightTime,8),
                  containGBKStrFormat(self.reachTime,8),
                   containGBKStrFormat(self.flightPort,16),
                   containGBKStrFormat(self.reachPort,16),
                   containGBKStrFormat(self.price,7),
                  containGBKStrFormat("转机" if self.isTransit else "不转机",7))
    @classmethod
    def tableHead(cls):
        return "%s%s%s%s%s%s%s%s%s" \
               % (containGBKStrFormat("订票ID", 10),
                  containGBKStrFormat("票源", 7),
                  containGBKStrFormat("航班", 16),
                  containGBKStrFormat("机型", 20),
                  containGBKStrFormat("起飞时间", 8),
                  containGBKStrFormat("到达时间", 8),
                  containGBKStrFormat("起飞机场", 16),
                  containGBKStrFormat("到达机场", 16),
                  containGBKStrFormat("价格", 7),
                  containGBKStrFormat("转机",7))
def containGBKStrFormat(strx, lenx):
    matchers=re.findall(r"[\u4e00-\u9fa5\u4dae\uE863]", strx)
    gbkStr=""
    for strv in matchers:
        gbkStr+=strv
    return '{x:^{y}s}'.format(x=strx,y=lenx - len(gbkStr.encode('GBK')) + len(gbkStr))
```

2. baseTicketWebsite类: 网站爬取操作的基类(定义了网站爬取及操作的通用方法)

def **init**(self,driver=None): 初始化一个网站操作,新建或复用一个可用的 无头的 selenium driver 浏览器

def getBeautifulSoup(self): 通过selenium driver 获取其beautifulSoup对象

def getHaveHeadChrome(self): 创建一个有头的 selenium driver 浏览器

def saveLoginCookie(self,url=None,path=None,driver=None): 保存某个url,driver的cookie信息

def readLoginCookie(self,url=None,path=None,driver=None): 读取某个url的cookie信息,传给driver

def bookTicketWindow(self): 创建一个有头的 订票 浏览器窗口

def logOut(self): 退出登录

```
import pickle

from bs4 import BeautifulSoup
from selenium.webdriver import Chrome
from selenium.webdriver import ChromeOptions
class baseTicketWebsite:
    def __init__(self,driver=None):
        if(not driver):
            option = ChromeOptions()
            option.add_experimental_option('excludeSwitches', ['enable-automation'])
            option.add_argument("--headless")
            prefs = {"profile.managed_default_content_settings.images": 2}
            #option.add_experimental_option("prefs", prefs)
            #option.add_argument("--no-sandbox")
            self.driver = Chrome(options=option)
            self.driver.execute_cdp_cmd("Page.addScriptToEvaluateOnNewDocument", {
                "source": """
                        Object.defineProperty(navigator, 'webdriver', {
                          get: () => undefined
                        })
                       """

            })
        else:
            self.driver=driver
    def getBeautifulSoup(self):
        return BeautifulSoup(self.driver.page_source, 'html.parser')
    def getHaveHeadChrome(self):
        option = ChromeOptions()
        option.add_experimental_option('excludeSwitches', ['enable-automation'])
        #option.add_argument("--headless")
        driver = Chrome(options=option)

        driver.execute_cdp_cmd("Page.addScriptToEvaluateOnNewDocument", {
            "source": """
                            Object.defineProperty(navigator, 'webdriver', {
                              get: () => undefined
                            })
                           """
        })
        return driver
```

```
    def driverget(self,url,isFlash=False):
        if self.driver.current_url==url:
            if isFlash:
                self.driver.get(url)
        else:
            self.driver.get(url)
    # 保存 某 driver的 url cookie 到 某个文件路径
    def saveLoginCookie(self,url=None,path=None,driver=None):
        if not driver:
            driver=self.driver
        if not url:
            url=self.url
        if not path:
            path=self.cookiePath
        if not (driver.current_url == url):
            driver.get(url)
        cookies=driver.get_cookies()
        with open(path,mode="wb") as filew:
            pickle.dump(cookies,filew)
    # 从某driver的 path 路径 读到 某个 url 的 cookie
    def readLoginCookie(self,url=None,path=None,driver=None):
        if not driver:
            driver=self.driver
        if not url:
            url=self.url
        if not path:
            path=self.cookiePath
        cookies=None

        try:
            if not (driver.current_url == url):
                driver.get(url)
            with open(path,mode="rb") as filer:
                cookies=pickle.load(file=filer)
                driver.delete_all_cookies()
                for cookie in cookies:
                    if "expiry" in cookie.keys():
                        # dict支持pop的删除函数
                        cookie.pop("expiry")
                    driver.add_cookie(cookie)
            driver.get(url)
            return True
        except FileNotFoundError as e :
            print("cookieNotFind")
            return False
    def bookTicketWindow(self):
        print("正在打开订票窗口,请注意桌面")
        driver = self.getHaveHeadChrome()
        self.readLoginCookie(driver=driver)
        driver.get(self.bookTicketUrl)
        self.otherDriver = driver
    def logOut(self):

        self.driver.delete_all_cookies()
```

```
        self.get(self.url)
```

3. ctrip类(继承baseTicketWebsite): 定义了携程网爬取及操作的方法

> def findAirline(self,fromCity,reachCity,fromDate):定义了如何根据 时间,地点去爬取携程网相应的机票信息
>
> def bookTicket(self,bookTicketId):定义了如何根据bookTicketId进行携程网订票窗口的打开
>
> def login(self):定义了如何登录携程网
>
> def loginHaveHeadWindow(self):定义了如何通过无头的窗口 让用户登录,并获取其登录的cookie信息
>
> def islogin(self):判断用户是否登录

```
import pickle
import re
import time
from random import random, randrange

from selenium.webdriver.common.keys import Keys
from selenium.webdriver.support.wait import WebDriverWait

#from airlineInfo import airlineInfo
#from ticketWebsite.baseTicketWebsite import baseTicketWebsite


class ctrip(baseTicketWebsite):
    name="携程网"
    url="https://flights.ctrip.com/international/search/domestic"
    cookiePath = "cookies/ctripCookie.txt"
    def __init__(self,driver=None):
        super(ctrip,self).__init__(driver)


    def findAirline(self,fromCity,reachCity,fromDate):

        airlineInfoList=[]
        self.driverget("https://flights.ctrip.com/international/search/domestic")
        fromCityE = WebDriverWait(self.driver, 20, poll_frequency=0.5,
ignored_exceptions=None) \
            .until(lambda dr: dr.find_element_by_css_selector("[name='owDCity']"))
        #fromCityE=self.driver.find_element_by_css_selector("[name='owDCity']")
        fromCityE.click()
        fromCityE.send_keys(fromCity)
        fromCityE.send_keys(Keys.ENTER)
        time.sleep(0.5)
        reachCityE=self.driver.find_element_by_css_selector("[name='owACity']")
        reachCityE.send_keys(Keys.CONTROL,'a')
        reachCityE.send_keys(Keys.BACKSPACE)
        reachCityE.send_keys(reachCity)
        time.sleep(0.2)

        reachCityE.send_keys(Keys.ENTER)
```

```
        self.driver.find_element_by_css_selector(".date-components input").click()
        time.sleep(0.5)
        fromCityId=re.search(r"[A-Z]+",fromCityE.get_attribute("value")).group().lower()
        reachCityId=re.search(r"[A-Z]+",reachCityE.get_attribute("value")).group().lower()
        self.tickInfoUrl =f"https://flights.ctrip.com/itinerary/oneway/{fromCityId}-
{reachCityId}?date={fromDate}&hasChild=false&hasBaby=false&classType=ALL"
        self.driverget(self.tickInfoUrl)

        element = WebDriverWait(self.driver,30,0.5).until(lambda
x:x.find_element_by_css_selector(".flight_card_content"))

 self.driver.execute_script("document.documentElement.scrollTop=document.documentElement.sc
rollHeight \n document.documentElement.scrollTop=document.documentElement.scrollHeight \n
document.documentElement.scrollTop=document.documentElement.scrollHeight")


 self.driver.execute_script("document.documentElement.scrollTop=document.documentElement.sc
rollHeight \n document.documentElement.scrollTop=document.documentElement.scrollHeight \n
document.documentElement.scrollTop=document.documentElement.scrollHeight")
        time.sleep(1)

 self.driver.execute_script("document.documentElement.scrollTop=document.documentElement.sc
rollHeight \n document.documentElement.scrollTop=document.documentElement.scrollHeight \n
document.documentElement.scrollTop=document.documentElement.scrollHeight")

        soup=self.getBeautifulSoup()
        divs=soup.select(".cabinV2 .Label_Flight")
        for fdiv in divs:
            airlineCompanyName=fdiv.select("[data-ubt-hover='c_flight_aircraftInfo']
strong")[0].text
            airlineId=fdiv.select("[data-ubt-hover='c_flight_aircraftInfo'] strong+span")
[0].string
            airlineName=airlineCompanyName+airlineId
            airlineModel=fdiv.select("[data-ubt-hover='c_flight_aircraftInfo']")
[1].select("span")[0].string
            flightTime=fdiv.select(".time")[0].string
            reachTime=fdiv.select(".time")[1].string
            flightPort=fdiv.select(".airport")[0].string
            reachPort=fdiv.select(".airport")[1].string
            flightOntimeRate=fdiv.select("[data-ubt-
hover='c_flight_punctualityRate_Flight'] span")[0].string if len(fdiv.select("[data-ubt-
hover='c_flight_punctualityRate_Flight'] span")) else ""
            price=fdiv.select(".base_price02")[0].text.replace("¥","")
            bookTicketId=airlineId

 airlineInfoList.append(airlineInfo(airlineName,airlineId,airlineModel,flightTime,

reachTime,flightPort,reachPort,flightOntimeRate,
                                           price,"携程",2,bookTicketId))
        divs = soup.select(".cabinV2 .Label_Transit")
        for fdiv in divs:

            airlineCompanyName = fdiv.select("[data-ubt-hover='c_flight_aircraftInfo']
```

```
strong")[0].text
            airlineId = fdiv.select("[data-ubt-hover='c_flight_aircraftInfo'] strong+span")
[0].string
            airlineName = airlineCompanyName + airlineId
            #airlineModel = fdiv.select("[data-ubt-hover='c_flight_aircraftInfo']")
[1].select("span")[0].string
            airlineModel = "暂无信息"
            flightTime = fdiv.select(".time")[0].string
            reachTime = fdiv.select(".time")[1].string
            flightPort = fdiv.select(".airport")[0].string
            reachPort = fdiv.select(".airport")[1].string
            #flightOntimeRate = fdiv.select("[data-ubt-
hover='c_flight_punctualityRate_Flight'] span")[0].string if len(
                #fdiv.select("[data-ubt-hover='c_flight_punctualityRate_Flight'] span"))
else ""
            price = fdiv.select(".base_price02")[0].text.replace("¥", "")
            bookTicketId = airlineId
            airlineInfoList.append(airlineInfo(airlineName, airlineId, airlineModel,
flightTime,
                                               reachTime, flightPort, reachPort,
flightOntimeRate,
                                               price, "携程", 2, bookTicketId,True))

        return airlineInfoList

    def bookTicket(self,bookTicketId):#btn_book arrow_down
        self.driverget(self.tickInfoUrl)
        self.driver.execute_script(
            "document.documentElement.scrollTop=document.documentElement.scrollHeight \n
document.documentElement.scrollTop=document.documentElement.scrollHeight \n
document.documentElement.scrollTop=document.documentElement.scrollHeight")

        self.driver.execute_script(
            "document.documentElement.scrollTop=document.documentElement.scrollHeight \n
document.documentElement.scrollTop=document.documentElement.scrollHeight \n
document.documentElement.scrollTop=document.documentElement.scrollHeight")
        time.sleep(2)
        self.driver.execute_script(
            "document.documentElement.scrollTop=document.documentElement.scrollHeight \n
document.documentElement.scrollTop=document.documentElement.scrollHeight \n
document.documentElement.scrollTop=document.documentElement.scrollHeight")
        if len(self.driver.find_elements_by_css_selector(".popup-tips .button")):
            self.driver.find_element_by_css_selector(".popup-tips .button").click()
        divs=self.driver.find_elements_by_css_selector(".cabinV2 .search_box")

        for divx in divs:
            textv=divx.find_element_by_css_selector("[data-ubt-
hover='c_flight_aircraftInfo'] strong+span").text
            #print(textv,bookTicketId)
            if textv==bookTicketId:
                xYJson = divx.location
                jsStr = "scrollTo(%d,%d)" % (xYJson["x"], xYJson["y"] - 100)

                self.driver.execute_script(jsStr)
```

```
                divx.find_element_by_css_selector(".btn_book.arrow_down").click()
                time.sleep(0.5)
                divx.find_element_by_css_selector(".inb.operations").click()
                time.sleep(0.5)
                self.bookTicketUrl = self.driver.current_url
                self.saveLoginCookie(self.bookTicketUrl)
                self.bookTicketWindow()
                break

    def login(self):
        self.readLoginCookie()
        text=self.islogin()

        if not text:
            self.loginHaveHeadWindow()
            self.readLoginCookie()
        text = self.islogin()
        print(f"{text}已经登录")
        return self
    def loginHaveHeadWindow(self):
        driverx = self.getHaveHeadChrome()
        driverx.get("https://passport.ctrip.com/user/login")

        #driver.execute_script("")
        #loginElement=driver.find_element_by_css_selector(".set-list.set-logIn a.person-
text")
        textD = WebDriverWait(driverx, 360, poll_frequency=1, ignored_exceptions=None) \
            .until(lambda dr: dr.find_element_by_css_selector(".member-name"))
        self.saveLoginCookie(driver=driverx)

        driverx.quit()

        return True

    def islogin(self):
        self.driver.get(self.url)
        textx=self.driver.find_element_by_css_selector(".member-name").text
        if(textx=="您好"):
            return None
        else:
            return "携程会员"+str(textx)
    def closeDriver(self):
        self.driver.quit()
```

4. fliggy类(继承baseTicketWebsite): 定义了飞猪网爬取及操作的方法

> def findAirline(self,fromCity,reachCity,fromDate):定义了如何根据 时间,地点去爬取飞猪网相应的机票信息
>
> def bookTicket(self,bookTicketId):定义了如何根据bookTicketId进行飞猪网订票窗口的打开
>
> def login(self):定义了如何登录飞猪网
>
> def loginOperation(self,username,password): 定义了如何使用账户密码直接登录飞猪网

def loginHaveHeadWindow(self):定义了如何通过无头的窗口 让用户登录,并获取其登录的cookie信息

def islogin(self):判断用户是否登录

```
import pickle
import re
import time
from random import randrange, random

#from selenium.webdriver import ActionChains
#from selenium.webdriver.support.wait import WebDriverWait

from airlineInfo import airlineInfo
from ticketWebsite.baseTicketWebsite import baseTicketWebsite


class fliggy(baseTicketWebsite):
    name="飞猪网"
    url="https://www.fliggy.com"
    cookiePath = "cookies/fliggyCookie.txt"
    def __init__(self,driver=None):
        super(fliggy,self).__init__(driver)


    def findAirline(self,fromCity,reachCity,fromDate):
        airlineInfoList=[]
        self.tickInfoUrl=f"https://sjipiao.fliggy.com/flight_search_result.htm?
_input_charset=utf-8&depCityName={fromCity}&depDate={fromDate}&arrCityName={reachCity}"
        self.driverget(self.tickInfoUrl)
        time.sleep(0.5)
        WebDriverWait(self.driver, 20, poll_frequency=0.5, ignored_exceptions=None) \
            .until(lambda dr: dr.find_element_by_css_selector(".flight-list-
item.clearfix.J_FlightItem"))
        soup=self.getBeautifulSoup()
        divs=soup.select(".flight-list-item.clearfix.J_FlightItem")
        for fdiv in divs:

            airlineName=fdiv.select(".J_line.J_TestFlight")[0].string
            airlineId=re.search(r"[0-9A-Z]+",airlineName).group()
            airlineModel=fdiv.select(".link-dot.J_FlightType")[0].string
            flightTime=fdiv.select(".flight-time-deptime")[0].string
            reachTime=fdiv.select(".s-time")[0].string
            flightPort=fdiv.select(".port-dep")[0].string
            reachPort=fdiv.select(".port-arr")[0].string
            flightOntimeRate=fdiv.select(".flight-ontime-rate")[0].string
            price=fdiv.select(".J_FlightListPrice")[0].string
            bookTicketId=airlineName

 airlineInfoList.append(airlineInfo(airlineName,airlineId,airlineModel,flightTime,

reachTime,flightPort,reachPort,flightOntimeRate,
                                                price,"飞猪网",1,bookTicketId))
```

```
        return airlineInfoList

    def bookTicket(self,bookTicketId):
        self.driverget(self.tickInfoUrl)
        divs=self.driver.find_elements_by_css_selector(".flight-list-
item.clearfix.J_FlightItem")
        for divx in divs:
            if
divx.find_element_by_css_selector(".J_line.J_TestFlight").text==bookTicketId:
                divx.find_element_by_css_selector(".select-btn.J_SelectFlight").click()
                time.sleep(0.5)
                bookAgents=self.driver.find_elements_by_css_selector(".agent-list-item")
                for bookAgent in bookAgents:
                    if(len(bookAgent.find_elements_by_css_selector("span[title='飞猪特
卖']"))):
                        bookAgent.find_element_by_css_selector(".select-
btn.J_Reserve").click()
                        time.sleep(0.5)
                        self.bookTicketUrl=self.driver.current_url
                        self.mySaveLoginCookie()
                        self.bookTicketWindow()
                        break
                break


    def loginOperation(self,username,password):
        self.driver.find_element_by_css_selector("#J_TLoginInfoHd>a").click()
        userNameInput = WebDriverWait(self.driver, 10, poll_frequency=0.2,
ignored_exceptions=None) \
            .until(lambda dr: dr.find_element_by_css_selector("#fm-login-id"))
        userNameInput.send_keys(username)
        self.driver.find_element_by_css_selector("#fm-login-password") \
            .send_keys(password)
        self.driver.find_element_by_css_selector(".fm-button.fm-submit").click()
        textD = WebDriverWait(self.driver, 10, poll_frequency=0.5, ignored_exceptions=None)
\
            .until(lambda dr: dr.find_element_by_css_selector(".userNickRegion-top"))

        return textD.text
    def login(self,username,password):
        self.myReadLoginCookie()
        text=self.islogin()

        if not text:
            try:
                text= self.loginOperation(username, password)
            except Exception as e:
                text=self.loginHaveHeadWindow()

        #text = self.islogin()
        print(f"{text}已经登录")
        self.mySaveLoginCookie()

        return self
```

```
    def loginHaveHeadWindow(self):
        driverx = self.getHaveHeadChrome()
        driverx.get(self.url)
        driverx.find_element_by_css_selector("#J_TLoginInfoHd>a").click()




        textD = WebDriverWait(driverx, 360, poll_frequency=0.5, ignored_exceptions=None) \
                    .until(lambda dr: dr.find_element_by_css_selector(".userNickRegion-
top"))

        self.saveLoginCookie(driver=driverx)
        self.saveLoginCookie(url="https://login.taobao.com/",driver=driverx)


        driverx.quit()
        self.readLoginCookie(url="https://login.taobao.com/")
        self.readsaveLoginCookie()


        return True

    def islogin(self):
        self.driverget(self.url,True)
        elements=self.driver.find_elements_by_css_selector(".userNickRegion-top")
        if len(elements):
            return elements[0].text
        else:
            return None
    def closeDriver(self):
        self.driver.quit()
    def mySaveLoginCookie(self):

        cookies=self.driver.get_cookies()
        with open("cookies/fliggyCookie.txt",mode="wb") as filew:
            pickle.dump(cookies,filew)
        self.driverget("https://www.taobao.com/")
        cookies = self.driver.get_cookies()
        with open("cookies/taobao.txt", mode="wb") as filew:
            pickle.dump(cookies, filew)
    def myReadLoginCookie(self):
        cookies=None

        try:
            self.driverget("https://www.taobao.com/")

            with open("cookies/taobao.txt",mode="rb") as filer:
                cookies=pickle.load(file=filer)
                self.driver.delete_all_cookies()
                for cookie in cookies:
                    if "expiry" in cookie.keys():

                        # dict支持pop的删除函数
```

```
                        cookie.pop("expiry")
                    self.driver.add_cookie(cookie)
            self.driverget(self.url)
            with open("cookies/fliggyCookie.txt",mode="rb") as filer:
                cookies=pickle.load(file=filer)
                self.driver.delete_all_cookies()
                for cookie in cookies:
                    if "expiry" in cookie.keys():
                        # dict支持pop的删除函数
                        cookie.pop("expiry")
                    self.driver.add_cookie(cookie)
            return True
        except FileNotFoundError as e :
            print("cookieNotFind")
            return False
    def bookTicketWindow(self):
        driver = self.getHaveHeadChrome()
        self.myReadLoginCookie(driver=driver)
        driver.get(self.bookTicketUrl)
        self.otherDriver = driver
```

5. qunar类(继承baseTicketWebsite): 定义了去哪儿网爬取及操作的方法

def findAirline(self,fromCity,reachCity,fromDate):定义了如何根据 时间,地点去爬取去哪儿网相应的信息

def bookTicket(self,bookTicketId):定义了如何根据bookTicketId进行去哪网订票窗口的打开

def login(self):定义了如何登录去哪网

def loginHaveHeadWindow(self):定义了如何通过无头的窗口 让用户登录,并获取其登录的cookie信息

def islogin(self):判断用户是否登录

```
import _thread
import pickle
import re
import time

from selenium.webdriver.common.keys import Keys
from selenium.webdriver.support.wait import WebDriverWait

#from airlineInfo import airlineInfo
#from ticketWebsite.baseTicketWebsite import baseTicketWebsite


class qunar(baseTicketWebsite):
    name="去哪网"
    url="https://flight.qunar.com"
    cookiePath="cookies/qunarCookie.txt"
    def __init__(self,driver=None):

        super(qunar,self).__init__(driver)
```

```
    def findAirline(self,fromCity,reachCity,fromDate):
        self.tickInfoUrl=f"https://flight.qunar.com/site/oneway_list.htm?
searchDepartureAirport={fromCity}&searchArrivalAirport={reachCity}&searchDepartureTime=
{fromDate}"

        self.driverget(self.tickInfoUrl)

        airlineInfoList = []
        while True :
            time.sleep(0.5)
            WebDriverWait(self.driver, 20, poll_frequency=0.5, ignored_exceptions=None) \
                 .until(lambda dr: dr.find_element_by_css_selector(".m-airfly-lst .b-
airfly"))
            soup=self.getBeautifulSoup()
            divs=soup.select(".m-airfly-lst .b-airfly")
            for fdiv in divs:

                divxx=fdiv.select(".air span")
                airlineName = divxx[0].string
                airlineId = fdiv.select(".num span")[0].string
                airlineModel = fdiv.select(".num span")[1].string
                isTransit=False
                if(len(fdiv.select(".num span"))>=4):
                    isTransit = True
                    transitAirName = divxx[0].string if len(divxx)==1 else divxx[1].string
                    transitLineId = fdiv.select(".num span")[2].string
                    transitAirModel = fdiv.select(".num span")[3].string

                flightTime=fdiv.select(".sep-lf>h2")[0].string
                reachTime=fdiv.select(".sep-rt>h2")[0].string
                flightPort=fdiv.select(".sep-lf .airport span")[0].string
                reachPort=fdiv.select(".sep-rt .airport span")[0].string
                flightOntimeRate=None
                priceS=fdiv.select(".prc_wp i")
                priceLen=len(priceS)

                price=""
                for i,p in enumerate(priceS):
                    fugaiPrice=fdiv.select(f"b[style*='left:-{(priceLen-i)*16}']")
                    if  not (len(fugaiPrice)):
                        price += p.string
                    else:
                        if(i==0 and len(fugaiPrice)==1):
                            price += p.string
                        else:
                            price += fugaiPrice[len(fugaiPrice)-1].string


                bookTicketId=airlineId

 airlineInfoList.append(airlineInfo(airlineName,airlineId,airlineModel,flightTime,
```

```
reachTime,flightPort,reachPort,flightOntimeRate,
                                                price,"去哪网",3,bookTicketId,
                                                isTransit,transitLineId if isTransit
else None,transitAirName if isTransit else None,transitAirModel if isTransit else None))
            nextPages=self.driver.find_elements_by_css_selector(".page-link")
            if  (not len(nextPages) ) or ("disabled" in
nextPages[1].get_attribute("class")):
                break
            else:
                nextPages[1].click()

        return airlineInfoList


    def bookTicket(self,bookTicketId):
        self.driverget(self.tickInfoUrl,True)
        while True:
            #time.sleep(0.5)
            WebDriverWait(self.driver, 10, poll_frequency=0.5, ignored_exceptions=None) \
                .until(lambda dr: dr.find_element_by_css_selector(".m-airfly-lst .b-
airfly"))
            divs = self.driver.find_elements_by_css_selector(".m-airfly-lst .b-airfly")
            for fdiv in divs:


                if(bookTicketId == fdiv.find_element_by_css_selector(".num span").text):
                    fdiv.find_element_by_css_selector(".num span").click()
                    WebDriverWait(self.driver, 10, poll_frequency=0.5,
ignored_exceptions=None) \
                        .until(lambda dr: dr.find_element_by_css_selector(".btn-book
")).click()
                    time.sleep(0.5)
                    for wh in self.driver.window_handles:
                        if(wh != self.driver.current_window_handle):
                            self.driver.switch_to.window(wh)
                            break
                    WebDriverWait(self.driver, 10, poll_frequency=0.5,
ignored_exceptions=None) \
                        .until(lambda dr: dr.find_element_by_css_selector(".section-ft .row
.button-wrap"))
                    self.bookTicketUrl = self.driver.current_url
                    self.saveLoginCookie(url=self.bookTicketUrl)

                    self.bookTicketWindow()
                    break

            nextPages = self.driver.find_elements_by_css_selector(".page-link")
            if (not len(nextPages)) or ("disabled" in nextPages[1].get_attribute("class")):
                break
            else:
                nextPages[1].click()
    def login(self):

        self.readLoginCookie()
```

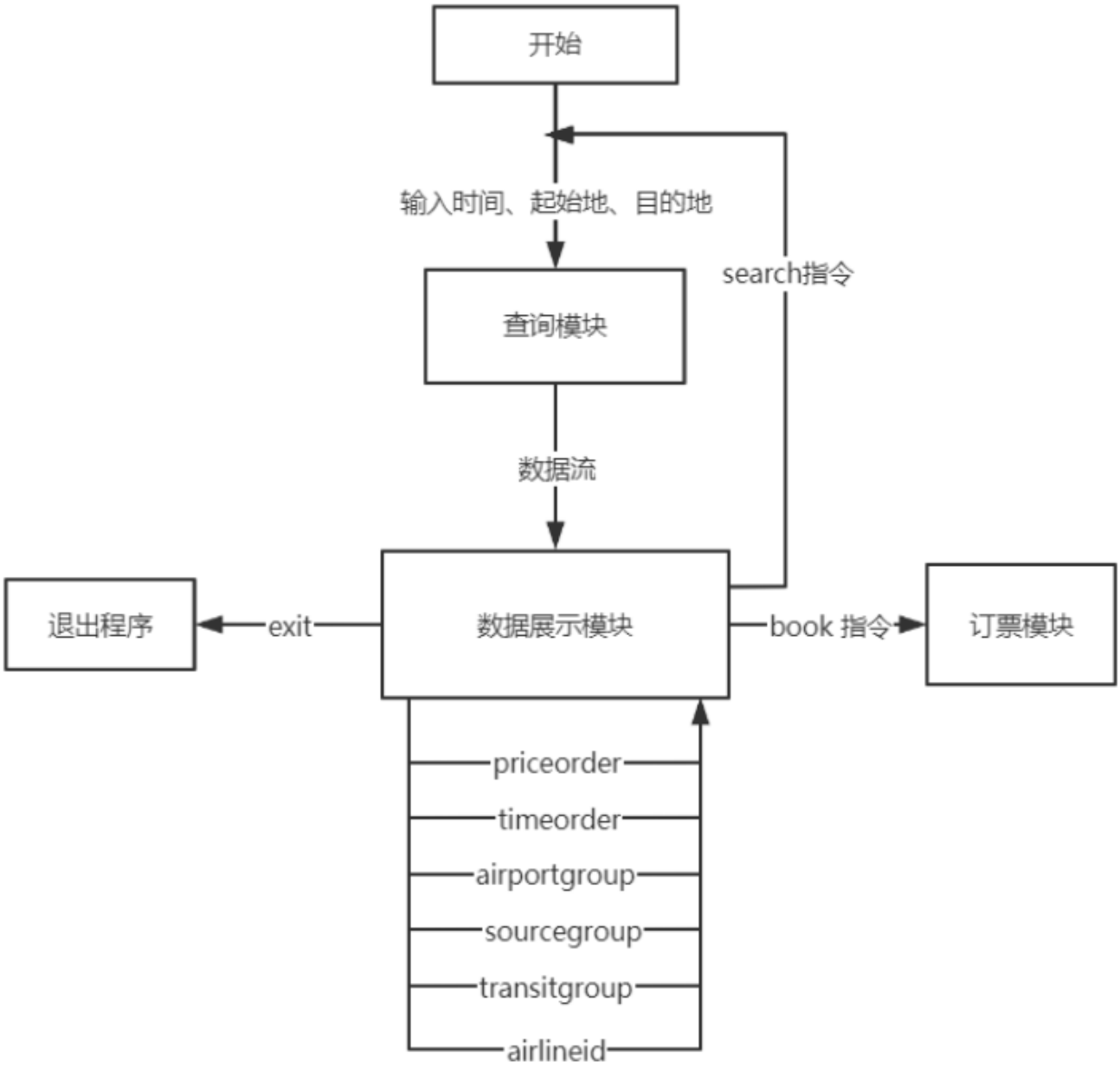
```
        text=self.islogin()

        if not text:
            text=self.loginHaveHeadWindow()
            self.readLoginCookie()
        text = self.islogin()
        print(f"{text}已经登录")
        return self
    def loginHaveHeadWindow(self):
        driver = self.getHaveHeadChrome()
        driver.get(self.url)
        driver.find_element_by_css_selector("#__headerInfo_login__").click()
        logName = WebDriverWait(driver, 360, poll_frequency=1, ignored_exceptions=None) \
            .until(lambda dr: dr.find_element_by_css_selector(".q_header_uname"))
        self.saveLoginCookie(driver=driver)
        driver.quit()
        return logName



    def islogin(self):
        self.driverget(self.url,True)
        elements=self.driver.find_elements_by_css_selector(".q_header_uname")
        if len(elements):
            return elements[0].text
        else:
            return None
    def closeDriver(self):
        self.driver.quit()
```

6. 程序入口: 定义了各类、模块之间的交互




```
import sys
import time


print("程序启动中")
fliggyO=fliggy().login("****淘宝账户名****","*****淘宝账户密码*****")
ctripO=ctrip(fliggyO.driver).login()
qunar0=qunar(ctripO.driver).login()
while True:
    print("*"*100)
    print("请输入查询航班的出发地")
    fromPort=input()
    print("*"*100)
    print("请输入查询航班的目的地")
    reachPort=input()

    print("*"*100)
```

```
print("请以YYYY-MM-DD格式输入要查询航班的日期")
fdate=input()
print("*"*100)
lista=[]
try:
    lista=fliggyO.findAirline(fromPort,reachPort,fdate)
except Exception as e:
    import traceback
    traceback.print_exc()
    print("飞猪发生错误")
    lista=[]
try:
    listb=ctripO.findAirline(fromPort,reachPort,fdate)
except Exception as e:
    import traceback
    traceback.print_exc()
    print("携程发生错误")
    listb=[]
try:
    listc=qunar0.findAirline(fromPort,reachPort,fdate)
except Exception as e:
    import traceback
    traceback.print_exc()
    print("去哪发生错误")
    listc = []
liste=lista+listb+listc
print(airlineInfo.tableHead())
liste.sort(key=lambda e:int(e.price))
for v in liste:
    print(v)
print("查询完毕")
while True:
    print("*"*100)
    inputStr=input("订票 book订票id \n"
                   "重新查询 search \n"
                   "价格升序 priceorder\n"
                  "价格降序 priceRorder\n"
                  "起飞时间升序 timeorder\n"
                  "起飞时间降序序 timeRorder\n"
                  "起飞机场分类 airportgroup\n"
                   "票源分类 sourcegroup\n"
                   "转机不转机分类 transitgroup\n"
                   "如果要进行航班比价 请输入 airlineid航班号\n"
                   "如果要退出 请输入 exit\n"
                  +("*"*50)+"请输入命令"+("*"*50)+"\n")
    print("*"*50+"接受到命令"+"*"*50)
    if("book" in inputStr):
        print("未登陆网站可能需要登陆或cookie登陆(请关注弹出的chrome窗口)")
        if (inputStr[4] == '1'):
            # fliggyO.bookTicket(inputStr[5:len(inputStr)])
            pass
        elif (inputStr[4] == '2'):

            ctripO.login().bookTicket(inputStr[5:len(inputStr)])
```

```
            elif (inputStr[4] == '3'):
                qunar0.login().bookTicket(inputStr[5:len(inputStr)])
        elif("search" in inputStr):
            break
        elif("order" in inputStr or "group" in inputStr):
            print(airlineInfo.tableHead())
            if("priceorder" in inputStr):
                keyx=lambda e: int(e.price)
            elif("priceRorder" in inputStr):
                keyx=lambda e: -int(e.price)
            elif ("timeorder" in inputStr):
                keyx=lambda e: e.flightTime
            elif ("timeRorder" in inputStr):
                keyx=lambda e: -1*int(e.flightTime.replace(":",""))
            elif ("airportgroup" in inputStr):
                keyx=lambda e: e.flightPort
            elif ("sourcegroup" in inputStr):
                keyx=lambda e: e.websiteId
            elif ("transitgroup" in inputStr):
                keyx = lambda e: e.isTransit
            liste.sort(key=keyx)
            for v in liste:
                print(v)
            print("排序操作完毕")
        elif("airlineid" in inputStr):
            print(airlineInfo.tableHead())
            airlineIdx=inputStr[9:len(inputStr)]
            airlineIdxList=[]
            for v in liste:
                if  v.airlineId==airlineIdx:
                    airlineIdxList.append(v)

            airlineIdxList.sort(key=lambda e:int(e.price))
            for v in airlineIdxList:
                print(v)
            print("比价完毕")
        elif ("exit" in inputStr):
            sys.exit(0)
        else:
            print("不可识别的程序代码,请重试")
```

7. 程序运行交互实例

```
 程序启动中
**********已经登录
********已经登录
*******已经登录
********************************************************************************************************

请输入查询航班的出发地
```

```
北京
********************************************************************************************************
请输入查询航班的目的地
青岛
********************************************************************************************************
请以YYYY-MM-DD格式输入要查询航班的日期
2020-12-24
********************************************************************************************************
  订票ID        票源                航班                机型          起飞时间     到达时间          起飞
机场                  到达机场            价格       转机
 3MU6965     去哪网         东方航空MU6965           空客320(中)          07:20       08:50
大兴机场                  流亭机场             180    不转机
 3MU5693     去哪网         东方航空MU5693           空客320(中)          08:15       17:40
大兴机场                  流亭机场             409    转机
 3ZH2844     去哪网         深圳航空ZH2844           波音738(中)          22:20       23:55
首都机场                  流亭机场             499    转机
 1SC4658     飞猪网         山东航空SC4658              738               22:20       23:55
北京首都国际机场T3          流亭国际机场          500    不转机
 2SC4658      携程          山东航空SC4658          波音738(中型)          22:20       23:55
首都国际机场T3          流亭国际机场T1        500    不转机
 3SC4658     去哪网         山东航空SC4658           波音738(中)          22:20       23:55
首都机场                  流亭机场             500    不转机
 3ZH2839     去哪网         深圳航空ZH2839           波音738(中)          06:45       08:15
首都机场                  流亭机场             509    转机
 1SC4649     飞猪网         山东航空SC4649              738               06:45       08:15
北京首都国际机场T3          流亭国际机场          510    不转机
 1MU6965     飞猪网       中国东方航空MU6965           325               07:20       08:50       北
京大兴国际机场           流亭国际机场T2        510    不转机
 2SC4649      携程          山东航空SC4649          波音738(中型)          06:45       08:15
首都国际机场T3          流亭国际机场T1        510    不转机
 2MU6965      携程          东方航空MU6965          空客321(中型)          07:20       08:50
大兴国际机场           流亭国际机场T2        510    不转机
 3SC4649     去哪网         山东航空SC4649           波音738(中)          06:45       08:15
首都机场                  流亭机场             510    不转机
 3CZ6134     去哪网         南方航空CZ6134           空客320(中)          18:15       23:50
大兴机场                  流亭机场             518    转机
 3ZH1525     去哪网         深圳航空ZH1525           波音738(中)          20:15       22:00
首都机场                  流亭机场             534    转机
 3CA1525     去哪网         中国国航CA1525           波音738(中)          20:15       22:00
首都机场                  流亭机场             543    不转机
 1CA1525     飞猪网         中国国航CA1525              738               20:15       22:00
北京首都国际机场T3          流亭国际机场          550    不转机
 2CA1525      携程          中国国航CA1525          波音738(中型)          20:15       22:00
首都国际机场T3          流亭国际机场T1        550    不转机
 3ZH1569     去哪网         深圳航空ZH1569           空客320(中)          08:30       10:05
首都机场                  流亭机场             554    转机
 3CA1569     去哪网         中国国航CA1569           空客320(中)          08:30       10:05
首都机场                  流亭机场             563    不转机
 1CA1569     飞猪网         中国国航CA1569              320               08:30       10:05
北京首都国际机场T3          流亭国际机场          570    不转机
 2CA1569      携程          中国国航CA1569          空客320(中型)          08:30       10:05
首都国际机场T3          流亭国际机场T1        570    不转机

 3ZH4910     去哪网         深圳航空ZH4910           波音737(中)          17:35       23:50
```

大兴机场 流亭机场 582 转机
3ZH1575 去哪网 深圳航空ZH1575 空客320(中) 09:30 11:05
首都机场 流亭机场 604 转机
3CA1575 去哪网 中国国航CA1575 空客320(中) 09:30 11:05
首都机场 流亭机场 613 不转机
1CA1575 飞猪网 中国国航CA1575 320 09:30 11:05
北京首都国际机场T3 流亭国际机场 620 不转机
1SC4652 飞猪网 山东航空SC4652 738 09:45 11:20
北京首都国际机场T3 流亭国际机场 620 不转机
1SC4656 飞猪网 山东航空SC4656 738 11:15 13:00
北京首都国际机场T3 流亭国际机场 620 不转机
1SC4654 飞猪网 山东航空SC4654 738 13:35 15:10
北京首都国际机场T3 流亭国际机场 620 不转机
2CA1575 携程 中国国航CA1575 空客320(中型) 09:30 11:05
首都国际机场T3 流亭国际机场T1 620 不转机
2SC4652 携程 山东航空SC4652 波音738(中型) 09:45 11:20
首都国际机场T3 流亭国际机场T1 620 不转机
2SC4656 携程 山东航空SC4656 波音738(中型) 11:15 13:00
首都国际机场T3 流亭国际机场T1 620 不转机
2SC4654 携程 山东航空SC4654 波音738(中型) 13:35 15:10
首都国际机场T3 流亭国际机场T1 620 不转机
3SC4652 去哪网 山东航空SC4652 波音738(中) 09:45 11:20
首都机场 流亭机场 620 不转机
3SC4656 去哪网 山东航空SC4656 波音738(中) 11:15 13:00
首都机场 流亭机场 620 不转机
3SC4654 去哪网 山东航空SC4654 波音738(中) 13:35 15:10
首都机场 流亭机场 620 不转机
3ZH1635 去哪网 深圳航空ZH1635 空客330(大) 18:35 00:05
首都机场 流亭机场 671 转机
3ZH1635 去哪网 深圳航空ZH1635 空客330(大) 18:35 23:55
首都机场 流亭机场 674 转机
3CZ6102 去哪网 南方航空CZ6102 空客321(中) 11:20 17:15
大兴机场 流亭机场 684 转机
3CZ6102 去哪网 南方航空CZ6102 空客321(中) 11:20 16:50
大兴机场 流亭机场 718 转机
3MU5120 去哪网 东方航空MU5120 空客330(大) 17:00 22:55
首都机场 流亭机场 770 转机
3ZH1571 去哪网 深圳航空ZH1571 空客321(中) 17:55 19:30
首都机场 流亭机场 804 转机
3CA1571 去哪网 中国国航CA1571 空客321(中) 17:55 19:30
首都机场 流亭机场 808 不转机
1CA1571 飞猪网 中国国航CA1571 321 17:55 19:30
北京首都国际机场T3 流亭国际机场 820 不转机
2CA1571 携程 中国国航CA1571 空客321(中型) 17:55 19:30
首都国际机场T3 流亭国际机场T1 820 不转机
2CA4649 携程 中国国航CA4649 波音738(中型) 06:45 08:15
首都国际机场T3 流亭国际机场T1 870 不转机
2CA4652 携程 中国国航CA4652 波音738(中型) 09:45 11:20
首都国际机场T3 流亭国际机场T1 870 不转机
2CA4654 携程 中国国航CA4654 波音738(中型) 13:35 15:10
首都国际机场T3 流亭国际机场T1 870 不转机

2CA4658 携程 中国国航CA4658 波音738(中型) 22:20 23:55

```
首都国际机场T3          流亭国际机场T1        870    不转机
 2SC1569      携程          山东航空SC1569           空客320(中型)          08:30       10:05
首都国际机场T3          流亭国际机场T1        940    不转机
 2SC1575      携程          山东航空SC1575           空客320(中型)          09:30       11:05
首都国际机场T3          流亭国际机场T1        940    不转机
 2SC1571      携程          山东航空SC1571           空客321(中型)          17:55       19:30
首都国际机场T3          流亭国际机场T1        940    不转机
 2SC1525      携程          山东航空SC1525           波音738(中型)          20:15       22:00
首都国际机场T3          流亭国际机场T1        940    不转机
查询完毕
**********************************************************************************************
订票 book订票id
重新查询 search
价格升序 priceorder
价格降序 priceRorder
起飞时间升序 timeorder
起飞时间降序序 timeRorder
起飞机场分类 airportgroup
票源分类 sourcegroup
转机不转机分类 transitgroup
如果要进行航班比价 请输入 airlineid航班号
如果要退出 请输入 exit
****************************************请输入命令****************************************
priceRorder
****************************************接受到命令****************************************
  订票ID        票源                航班                机型          起飞时间     到达时间           起飞
机场                  到达机场             价格       转机
 2SC1569      携程          山东航空SC1569           空客320(中型)          08:30       10:05
首都国际机场T3          流亭国际机场T1        940    不转机
 2SC1575      携程          山东航空SC1575           空客320(中型)          09:30       11:05
首都国际机场T3          流亭国际机场T1        940    不转机
 2SC1571      携程          山东航空SC1571           空客321(中型)          17:55       19:30
首都国际机场T3          流亭国际机场T1        940    不转机
 2SC1525      携程          山东航空SC1525           波音738(中型)          20:15       22:00
首都国际机场T3          流亭国际机场T1        940    不转机
 2CA4649      携程          中国国航CA4649           波音738(中型)          06:45       08:15
首都国际机场T3          流亭国际机场T1        870    不转机
 2CA4652      携程          中国国航CA4652           波音738(中型)          09:45       11:20
首都国际机场T3          流亭国际机场T1        870    不转机
 2CA4654      携程          中国国航CA4654           波音738(中型)          13:35       15:10
首都国际机场T3          流亭国际机场T1        870    不转机
 2CA4658      携程          中国国航CA4658           波音738(中型)          22:20       23:55
首都国际机场T3          流亭国际机场T1        870    不转机
 1CA1571     飞猪网          中国国航CA1571                321               17:55       19:30
北京首都国际机场T3          流亭国际机场         820    不转机
 2CA1571      携程          中国国航CA1571           空客321(中型)          17:55       19:30
首都国际机场T3          流亭国际机场T1        820    不转机
 3CA1571     去哪网          中国国航CA1571             空客321(中)          17:55       19:30
首都机场                流亭机场              808    不转机
 3ZH1571     去哪网          深圳航空ZH1571             空客321(中)          17:55       19:30
首都机场                流亭机场              804     转机


 3MU5120     去哪网          东方航空MU5120             空客330(大)          17:00       22:55
```

首都机场 流亭机场 770 转机
3CZ6102 去哪网 南方航空CZ6102 空客321(中) 11:20 16:50
大兴机场 流亭机场 718 转机
3CZ6102 去哪网 南方航空CZ6102 空客321(中) 11:20 17:15
大兴机场 流亭机场 684 转机
3ZH1635 去哪网 深圳航空ZH1635 空客330(大) 18:35 23:55
首都机场 流亭机场 674 转机
3ZH1635 去哪网 深圳航空ZH1635 空客330(大) 18:35 00:05
首都机场 流亭机场 671 转机
1CA1575 飞猪网 中国国航CA1575 320 09:30 11:05
北京首都国际机场T3 流亭国际机场 620 不转机
1SC4652 飞猪网 山东航空SC4652 738 09:45 11:20
北京首都国际机场T3 流亭国际机场 620 不转机
1SC4656 飞猪网 山东航空SC4656 738 11:15 13:00
北京首都国际机场T3 流亭国际机场 620 不转机
1SC4654 飞猪网 山东航空SC4654 738 13:35 15:10
北京首都国际机场T3 流亭国际机场 620 不转机
2CA1575 携程 中国国航CA1575 空客320(中型) 09:30 11:05
首都国际机场T3 流亭国际机场T1 620 不转机
2SC4652 携程 山东航空SC4652 波音738(中型) 09:45 11:20
首都国际机场T3 流亭国际机场T1 620 不转机
2SC4656 携程 山东航空SC4656 波音738(中型) 11:15 13:00
首都国际机场T3 流亭国际机场T1 620 不转机
2SC4654 携程 山东航空SC4654 波音738(中型) 13:35 15:10
首都国际机场T3 流亭国际机场T1 620 不转机
3SC4652 去哪网 山东航空SC4652 波音738(中) 09:45 11:20
首都机场 流亭机场 620 不转机
3SC4656 去哪网 山东航空SC4656 波音738(中) 11:15 13:00
首都机场 流亭机场 620 不转机
3SC4654 去哪网 山东航空SC4654 波音738(中) 13:35 15:10
首都机场 流亭机场 620 不转机
3CA1575 去哪网 中国国航CA1575 空客320(中) 09:30 11:05
首都机场 流亭机场 613 不转机
3ZH1575 去哪网 深圳航空ZH1575 空客320(中) 09:30 11:05
首都机场 流亭机场 604 转机
3ZH4910 去哪网 深圳航空ZH4910 波音737(中) 17:35 23:50
大兴机场 流亭机场 582 转机
1CA1569 飞猪网 中国国航CA1569 320 08:30 10:05
北京首都国际机场T3 流亭国际机场 570 不转机
2CA1569 携程 中国国航CA1569 空客320(中型) 08:30 10:05
首都国际机场T3 流亭国际机场T1 570 不转机
3CA1569 去哪网 中国国航CA1569 空客320(中) 08:30 10:05
首都机场 流亭机场 563 不转机
3ZH1569 去哪网 深圳航空ZH1569 空客320(中) 08:30 10:05
首都机场 流亭机场 554 转机
1CA1525 飞猪网 中国国航CA1525 738 20:15 22:00
北京首都国际机场T3 流亭国际机场 550 不转机
2CA1525 携程 中国国航CA1525 波音738(中型) 20:15 22:00
首都国际机场T3 流亭国际机场T1 550 不转机
3CA1525 去哪网 中国国航CA1525 波音738(中) 20:15 22:00
首都机场 流亭机场 543 不转机

3ZH1525 去哪网 深圳航空ZH1525 波音738(中) 20:15 22:00

首都机场 流亭机场 534 转机
3CZ6134 去哪网 南方航空CZ6134 空客320(中) 18:15 23:50
大兴机场 流亭机场 518 转机
1SC4649 飞猪网 山东航空SC4649 738 06:45 08:15
北京首都国际机场T3 流亭国际机场 510 不转机
1MU6965 飞猪网 中国东方航空MU6965 325 07:20 08:50 北
京大兴国际机场 流亭国际机场T2 510 不转机
2SC4649 携程 山东航空SC4649 波音738(中型) 06:45 08:15
首都国际机场T3 流亭国际机场T1 510 不转机
2MU6965 携程 东方航空MU6965 空客321(中型) 07:20 08:50
大兴国际机场 流亭国际机场T2 510 不转机
3SC4649 去哪网 山东航空SC4649 波音738(中) 06:45 08:15
首都机场 流亭机场 510 不转机
3ZH2839 去哪网 深圳航空ZH2839 波音738(中) 06:45 08:15
首都机场 流亭机场 509 转机
1SC4658 飞猪网 山东航空SC4658 738 22:20 23:55
北京首都国际机场T3 流亭国际机场 500 不转机
2SC4658 携程 山东航空SC4658 波音738(中型) 22:20 23:55
首都国际机场T3 流亭国际机场T1 500 不转机
3SC4658 去哪网 山东航空SC4658 波音738(中) 22:20 23:55
首都机场 流亭机场 500 不转机
3ZH2844 去哪网 深圳航空ZH2844 波音738(中) 22:20 23:55
首都机场 流亭机场 499 转机
3MU5693 去哪网 东方航空MU5693 空客320(中) 08:15 17:40
大兴机场 流亭机场 409 转机
3MU6965 去哪网 东方航空MU6965 空客320(中) 07:20 08:50
大兴机场 流亭机场 180 不转机
排序操作完毕
****
订票 book订票id
重新查询 search
价格升序 priceorder
价格降序 priceRorder
起飞时间升序 timeorder
起飞时间降序序 timeRorder
起飞机场分类 airportgroup
票源分类 sourcegroup
转机不转机分类 transitgroup
如果要进行航班比价 请输入 airlineid航班号
如果要退出 请输入 exit
**********请输入命令**********
timeRorder
**********接受到命令**********
订票ID 票源 航班 机型 起飞时间 到达时间 起飞
机场 到达机场 价格 转机
2CA4658 携程 中国国航CA4658 波音738(中型) 22:20 23:55
首都国际机场T3 流亭国际机场T1 870 不转机
1SC4658 飞猪网 山东航空SC4658 738 22:20 23:55
北京首都国际机场T3 流亭国际机场 500 不转机
2SC4658 携程 山东航空SC4658 波音738(中型) 22:20 23:55
首都国际机场T3 流亭国际机场T1 500 不转机

3SC4658 去哪网 山东航空SC4658 波音738(中) 22:20 23:55

首都机场 流亭机场 500 不转机
3ZH2844 去哪网 深圳航空ZH2844 波音738(中) 22:20 23:55
首都机场 流亭机场 499 转机
2SC1525 携程 山东航空SC1525 波音738(中型) 20:15 22:00
首都国际机场T3 流亭国际机场T1 940 不转机
1CA1525 飞猪网 中国国航CA1525 738 20:15 22:00
北京首都国际机场T3 流亭国际机场 550 不转机
2CA1525 携程 中国国航CA1525 波音738(中型) 20:15 22:00
首都国际机场T3 流亭国际机场T1 550 不转机
3CA1525 去哪网 中国国航CA1525 波音738(中) 20:15 22:00
首都机场 流亭机场 543 不转机
3ZH1525 去哪网 深圳航空ZH1525 波音738(中) 20:15 22:00
首都机场 流亭机场 534 转机
3ZH1635 去哪网 深圳航空ZH1635 空客330(大) 18:35 23:55
首都机场 流亭机场 674 转机
3ZH1635 去哪网 深圳航空ZH1635 空客330(大) 18:35 00:05
首都机场 流亭机场 671 转机
3CZ6134 去哪网 南方航空CZ6134 空客320(中) 18:15 23:50
大兴机场 流亭机场 518 转机
2SC1571 携程 山东航空SC1571 空客321(中型) 17:55 19:30
首都国际机场T3 流亭国际机场T1 940 不转机
1CA1571 飞猪网 中国国航CA1571 321 17:55 19:30
北京首都国际机场T3 流亭国际机场 820 不转机
2CA1571 携程 中国国航CA1571 空客321(中型) 17:55 19:30
首都国际机场T3 流亭国际机场T1 820 不转机
3CA1571 去哪网 中国国航CA1571 空客321(中) 17:55 19:30
首都机场 流亭机场 808 不转机
3ZH1571 去哪网 深圳航空ZH1571 空客321(中) 17:55 19:30
首都机场 流亭机场 804 转机
3ZH4910 去哪网 深圳航空ZH4910 波音737(中) 17:35 23:50
大兴机场 流亭机场 582 转机
3MU5120 去哪网 东方航空MU5120 空客330(大) 17:00 22:55
首都机场 流亭机场 770 转机
2CA4654 携程 中国国航CA4654 波音738(中型) 13:35 15:10
首都国际机场T3 流亭国际机场T1 870 不转机
1SC4654 飞猪网 山东航空SC4654 738 13:35 15:10
北京首都国际机场T3 流亭国际机场 620 不转机
2SC4654 携程 山东航空SC4654 波音738(中型) 13:35 15:10
首都国际机场T3 流亭国际机场T1 620 不转机
3SC4654 去哪网 山东航空SC4654 波音738(中) 13:35 15:10
首都机场 流亭机场 620 不转机
3CZ6102 去哪网 南方航空CZ6102 空客321(中) 11:20 16:50
大兴机场 流亭机场 718 转机
3CZ6102 去哪网 南方航空CZ6102 空客321(中) 11:20 17:15
大兴机场 流亭机场 684 转机
1SC4656 飞猪网 山东航空SC4656 738 11:15 13:00
北京首都国际机场T3 流亭国际机场 620 不转机
2SC4656 携程 山东航空SC4656 波音738(中型) 11:15 13:00
首都国际机场T3 流亭国际机场T1 620 不转机
3SC4656 去哪网 山东航空SC4656 波音738(中) 11:15 13:00
首都机场 流亭机场 620 不转机

2CA4652 携程 中国国航CA4652 波音738(中型) 09:45 11:20

首都国际机场T3 流亭国际机场T1 870 不转机
1SC4652 飞猪网 山东航空SC4652 738 09:45 11:20
北京首都国际机场T3 流亭国际机场 620 不转机
2SC4652 携程 山东航空SC4652 波音738(中型) 09:45 11:20
首都国际机场T3 流亭国际机场T1 620 不转机
3SC4652 去哪网 山东航空SC4652 波音738(中) 09:45 11:20
首都机场 流亭机场 620 不转机
2SC1575 携程 山东航空SC1575 空客320(中型) 09:30 11:05
首都国际机场T3 流亭国际机场T1 940 不转机
1CA1575 飞猪网 中国国航CA1575 320 09:30 11:05
北京首都国际机场T3 流亭国际机场 620 不转机
2CA1575 携程 中国国航CA1575 空客320(中型) 09:30 11:05
首都国际机场T3 流亭国际机场T1 620 不转机
3CA1575 去哪网 中国国航CA1575 空客320(中) 09:30 11:05
首都机场 流亭机场 613 不转机
3ZH1575 去哪网 深圳航空ZH1575 空客320(中) 09:30 11:05
首都机场 流亭机场 604 转机
2SC1569 携程 山东航空SC1569 空客320(中型) 08:30 10:05
首都国际机场T3 流亭国际机场T1 940 不转机
1CA1569 飞猪网 中国国航CA1569 320 08:30 10:05
北京首都国际机场T3 流亭国际机场 570 不转机
2CA1569 携程 中国国航CA1569 空客320(中型) 08:30 10:05
首都国际机场T3 流亭国际机场T1 570 不转机
3CA1569 去哪网 中国国航CA1569 空客320(中) 08:30 10:05
首都机场 流亭机场 563 不转机
3ZH1569 去哪网 深圳航空ZH1569 空客320(中) 08:30 10:05
首都机场 流亭机场 554 转机
3MU5693 去哪网 东方航空MU5693 空客320(中) 08:15 17:40
大兴机场 流亭机场 409 转机
1MU6965 飞猪网 中国东方航空MU6965 325 07:20 08:50 北京大兴国际机场 流亭国际机场T2 510 不转机
2MU6965 携程 东方航空MU6965 空客321(中型) 07:20 08:50
大兴国际机场 流亭国际机场T2 510 不转机
3MU6965 去哪网 东方航空MU6965 空客320(中) 07:20 08:50
大兴机场 流亭机场 180 不转机
2CA4649 携程 中国国航CA4649 波音738(中型) 06:45 08:15
首都国际机场T3 流亭国际机场T1 870 不转机
1SC4649 飞猪网 山东航空SC4649 738 06:45 08:15
北京首都国际机场T3 流亭国际机场 510 不转机
2SC4649 携程 山东航空SC4649 波音738(中型) 06:45 08:15
首都国际机场T3 流亭国际机场T1 510 不转机
3SC4649 去哪网 山东航空SC4649 波音738(中) 06:45 08:15
首都机场 流亭机场 510 不转机
3ZH2839 去哪网 深圳航空ZH2839 波音738(中) 06:45 08:15
首都机场 流亭机场 509 转机
排序操作完毕
****
订票 book订票id
重新查询 search
价格升序 priceorder
价格降序 priceRorder
起飞时间升序 timeorder

起飞时间降序序 timeRorder

```
起飞机场分类 airportgroup
票源分类 sourcegroup
转机不转机分类 transitgroup
如果要进行航班比价 请输入 airlineid航班号
如果要退出 请输入 exit
****************************************请输入命令****************************************
sourcegroup
****************************************接受到命令****************************************
  订票ID        票源                航班                机型          起飞时间     到达时间          起飞
机场                  到达机场            价格       转机
 1SC4658     飞猪网        山东航空SC4658               738              22:20       23:55
北京首都国际机场T3          流亭国际机场         500   不转机
 1CA1525     飞猪网        中国国航CA1525               738              20:15       22:00
北京首都国际机场T3          流亭国际机场         550   不转机
 1CA1571     飞猪网        中国国航CA1571               321              17:55       19:30
北京首都国际机场T3          流亭国际机场         820   不转机
 1SC4654     飞猪网        山东航空SC4654               738              13:35       15:10
北京首都国际机场T3          流亭国际机场         620   不转机
 1SC4656     飞猪网        山东航空SC4656               738              11:15       13:00
北京首都国际机场T3          流亭国际机场         620   不转机
 1SC4652     飞猪网        山东航空SC4652               738              09:45       11:20
北京首都国际机场T3          流亭国际机场         620   不转机
 1CA1575     飞猪网        中国国航CA1575               320              09:30       11:05
北京首都国际机场T3          流亭国际机场         620   不转机
 1CA1569     飞猪网        中国国航CA1569               320              08:30       10:05
北京首都国际机场T3          流亭国际机场         570   不转机
 1MU6965     飞猪网      中国东方航空MU6965          325              07:20       08:50       北
京大兴国际机场          流亭国际机场T2       510   不转机
 1SC4649     飞猪网        山东航空SC4649               738              06:45       08:15
北京首都国际机场T3          流亭国际机场         510   不转机
 2CA4658      携程         中国国航CA4658          波音738(中型)          22:20       23:55
首都国际机场T3         流亭国际机场T1       870   不转机
 2SC4658      携程         山东航空SC4658          波音738(中型)          22:20       23:55
首都国际机场T3         流亭国际机场T1       500   不转机
 2SC1525      携程         山东航空SC1525          波音738(中型)          20:15       22:00
首都国际机场T3         流亭国际机场T1       940   不转机
 2CA1525      携程         中国国航CA1525          波音738(中型)          20:15       22:00
首都国际机场T3         流亭国际机场T1       550   不转机
 2SC1571      携程         山东航空SC1571          空客321(中型)          17:55       19:30
首都国际机场T3         流亭国际机场T1       940   不转机
 2CA1571      携程         中国国航CA1571          空客321(中型)          17:55       19:30
首都国际机场T3         流亭国际机场T1       820   不转机
 2CA4654      携程         中国国航CA4654          波音738(中型)          13:35       15:10
首都国际机场T3         流亭国际机场T1       870   不转机
 2SC4654      携程         山东航空SC4654          波音738(中型)          13:35       15:10
首都国际机场T3         流亭国际机场T1       620   不转机
 2SC4656      携程         山东航空SC4656          波音738(中型)          11:15       13:00
首都国际机场T3         流亭国际机场T1       620   不转机
 2CA4652      携程         中国国航CA4652          波音738(中型)          09:45       11:20
首都国际机场T3         流亭国际机场T1       870   不转机
 2SC4652      携程         山东航空SC4652          波音738(中型)          09:45       11:20
首都国际机场T3         流亭国际机场T1       620   不转机

 2SC1575      携程         山东航空SC1575          空客320(中型)          09:30       11:05
```

首都国际机场T3 流亭国际机场T1 940 不转机
2CA1575 携程 中国国航CA1575 空客320(中型) 09:30 11:05
首都国际机场T3 流亭国际机场T1 620 不转机
2SC1569 携程 山东航空SC1569 空客320(中型) 08:30 10:05
首都国际机场T3 流亭国际机场T1 940 不转机
2CA1569 携程 中国国航CA1569 空客320(中型) 08:30 10:05
首都国际机场T3 流亭国际机场T1 570 不转机
2MU6965 携程 东方航空MU6965 空客321(中型) 07:20 08:50
大兴国际机场 流亭国际机场T2 510 不转机
2CA4649 携程 中国国航CA4649 波音738(中型) 06:45 08:15
首都国际机场T3 流亭国际机场T1 870 不转机
2SC4649 携程 山东航空SC4649 波音738(中型) 06:45 08:15
首都国际机场T3 流亭国际机场T1 510 不转机
3SC4658 去哪网 山东航空SC4658 波音738(中) 22:20 23:55
首都机场 流亭机场 500 不转机
3ZH2844 去哪网 深圳航空ZH2844 波音738(中) 22:20 23:55
首都机场 流亭机场 499 转机
3CA1525 去哪网 中国国航CA1525 波音738(中) 20:15 22:00
首都机场 流亭机场 543 不转机
3ZH1525 去哪网 深圳航空ZH1525 波音738(中) 20:15 22:00
首都机场 流亭机场 534 转机
3ZH1635 去哪网 深圳航空ZH1635 空客330(大) 18:35 23:55
首都机场 流亭机场 674 转机
3ZH1635 去哪网 深圳航空ZH1635 空客330(大) 18:35 00:05
首都机场 流亭机场 671 转机
3CZ6134 去哪网 南方航空CZ6134 空客320(中) 18:15 23:50
大兴机场 流亭机场 518 转机
3CA1571 去哪网 中国国航CA1571 空客321(中) 17:55 19:30
首都机场 流亭机场 808 不转机
3ZH1571 去哪网 深圳航空ZH1571 空客321(中) 17:55 19:30
首都机场 流亭机场 804 转机
3ZH4910 去哪网 深圳航空ZH4910 波音737(中) 17:35 23:50
大兴机场 流亭机场 582 转机
3MU5120 去哪网 东方航空MU5120 空客330(大) 17:00 22:55
首都机场 流亭机场 770 转机
3SC4654 去哪网 山东航空SC4654 波音738(中) 13:35 15:10
首都机场 流亭机场 620 不转机
3CZ6102 去哪网 南方航空CZ6102 空客321(中) 11:20 16:50
大兴机场 流亭机场 718 转机
3CZ6102 去哪网 南方航空CZ6102 空客321(中) 11:20 17:15
大兴机场 流亭机场 684 转机
3SC4656 去哪网 山东航空SC4656 波音738(中) 11:15 13:00
首都机场 流亭机场 620 不转机
3SC4652 去哪网 山东航空SC4652 波音738(中) 09:45 11:20
首都机场 流亭机场 620 不转机
3CA1575 去哪网 中国国航CA1575 空客320(中) 09:30 11:05
首都机场 流亭机场 613 不转机
3ZH1575 去哪网 深圳航空ZH1575 空客320(中) 09:30 11:05
首都机场 流亭机场 604 转机
3CA1569 去哪网 中国国航CA1569 空客320(中) 08:30 10:05
首都机场 流亭机场 563 不转机

3ZH1569 去哪网 深圳航空ZH1569 空客320(中) 08:30 10:05

```
首都机场                流亭机场            554    转机
 3MU5693    去哪网        东方航空MU5693          空客320(中)          08:15       17:40
大兴机场                流亭机场            409    转机
 3MU6965    去哪网        东方航空MU6965          空客320(中)          07:20       08:50
大兴机场                流亭机场            180   不转机
 3SC4649    去哪网        山东航空SC4649          波音738(中)          06:45       08:15
首都机场                流亭机场            510   不转机
 3ZH2839    去哪网        深圳航空ZH2839          波音738(中)          06:45       08:15
首都机场                流亭机场            509    转机
排序操作完毕
**********************************************************************************************
订票 book订票id
重新查询 search
价格升序 priceorder
价格降序 priceRorder
起飞时间升序 timeorder
起飞时间降序序 timeRorder
起飞机场分类 airportgroup
票源分类 sourcegroup
转机不转机分类 transitgroup
如果要进行航班比价 请输入 airlineid航班号
如果要退出 请输入 exit
****************************************请输入命令****************************************
airlineidMU6965
****************************************接受到命令****************************************
  订票ID       票源              航班                机型          起飞时间     到达时间          起飞
机场                到达机场            价格       转机
 3MU6965    去哪网        东方航空MU6965          空客320(中)          07:20       08:50
大兴机场                流亭机场            180   不转机
 1MU6965    飞猪网       中国东方航空MU6965           325              07:20       08:50       北
京大兴国际机场          流亭国际机场T2       510   不转机
 2MU6965     携程         东方航空MU6965          空客321(中型)          07:20       08:50
大兴国际机场          流亭国际机场T2       510   不转机
比价完毕
**********************************************************************************************
****
订票 book订票id
重新查询 search
价格升序 priceorder
价格降序 priceRorder
起飞时间升序 timeorder
起飞时间降序序 timeRorder
起飞机场分类 airportgroup
票源分类 sourcegroup
转机不转机分类 transitgroup
如果要进行航班比价 请输入 airlineid航班号
如果要退出 请输入 exit
****************************************请输入命令****************************************
book3MU6965
****************************************接受到命令****************************************
未登陆网站可能需要登陆或cookie登陆(请关注弹出的chrome窗口)
cikn8049已经登录

订票窗口已弹出,请进行信息补充
```

```
**********************************************************************************************
订票 book订票id
重新查询 search
价格升序 priceorder
价格降序 priceRorder
起飞时间升序 timeorder
起飞时间降序序 timeRorder
起飞机场分类 airportgroup
票源分类 sourcegroup
转机不转机分类 transitgroup
如果要进行航班比价 请输入 airlineid航班号
如果要退出 请输入 exit
****************************************请输入命令****************************************
book2MU6965
****************************************接受到命令****************************************
未登陆网站可能需要登陆或cookie登陆(请关注弹出的chrome窗口)
携程会员已经登录
订票窗口已弹出,请进行信息补充
**********************************************************************************************
订票 book订票id
重新查询 search
价格升序 priceorder
价格降序 priceRorder
起飞时间升序 timeorder
起飞时间降序序 timeRorder
起飞机场分类 airportgroup
票源分类 sourcegroup
转机不转机分类 transitgroup
如果要进行航班比价 请输入 airlineid航班号
如果要退出 请输入 exit
****************************************请输入命令****************************************
exit
****************************************接受到命令****************************************
An exception has occurred, use %tb to see the full traceback.
SystemExit: 0
```

结论

> 在这个互联网信息爆炸的时代，有时候单纯的依靠浏览网页，已经无法满足我们获取信息质量和数量的要求，但厂商的数据保护与反爬虫意识日益增强，导致我们爬取数据的难度也日益增加。这次的爬虫作业，我就遇到了很多超出预期的困难和挑战，像淘宝飞猪网对selenium的识别和限制，以及去哪网对机票价格的混淆等等，虽然最终都解决了问题, 但感觉在一些细节上仍然需要继续学习深入。当然爬虫作为一门与时俱进的技术 ,在我们已经掌握现有基础的前提下,更注重的则是实战的应变能力，不同网站的反爬策略可能不一样，我们在扎实基础的前提上有时候也要灵活应变，才能获取到我们想要的结果。

参考资料与文献

1: 单点登录，第三方cookie和cna

2: 淘宝反扒解决方案