ZE抓数步骤:

流程描述:

1. 先向首页发送请求,获得一个用于后面请求航班的参数,该参数有效期至少4小时
2. 拿该参数请求获得数据
3. 获得数据,解析航班

具体实现:

步骤1:

请求地址: **https://www.eastarjet.com/newstar/PGWHC00001**

请求头:**headers = {**

**"Host": "www.eastarjet.com"**

**,"Connection": "keep-alive"**

**,"Upgrade-Insecure-Requests": "1"**

**,"User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/77.0.3865.90 Safari/537.36"**

**,"Sec-Fetch-Mode": "navigate"**

**,"Sec-Fetch-User": "?1"**

**,"Accept": "text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,image/apng,\*/\*;q=0.8,application/signed-exchange;v=b3"**

**,"Sec-Fetch-Site": "none"**

**,"Accept-Encoding": "gzip, deflate, br"**

**,"Accept-Language": "zh-CN,zh;q=0.9"**

**}**

请求方式:get

示例:响应头

**{'Access-Control-Allow-Origin': '\*', 'cache n$cu=1570590792582;Path=/;Max-Age=30, n$cu=1570590792583;Path=/;Max-Age=30, n$cu=15705907925-control': 'no-store, no-cache, must-revalith=/;Max-Age=30, n$cu=1570590792583;Path=/;Max-Age=30, n$cu=1570590792583;Path=/;Max-Age=30,date', 'Content-Type': 'text/html;charset=U1570590792583;Path=/;Max-Age=30, selected\_culture\_code=zh\_CN; Expires=Thu, 08-Oct-2020 03:13:TF-8', 'Date': 'Wed, 09 Oct 2019 03:13:12 Gxpires=Thu, 08-Oct-2020 03:13:12 GMT; Path=/, JSESSIONID=8FD62FD4FA71046A76E5C47D956D7BFB.WASMT', 'expires': '0', 'pragma': 'no-cache', .WAS\_213; path=/;', 'transfer-encoding': 'chunked', 'Connection': 'keep-alive'}**

**'Server': 'Apache-Coyote/1.1', 'Set-Cookie': 'n$D=1;Path=/;Max-Age=86400, n$H=1;Path=/>;Max-Age=3600, n$cu=1570590792582;Path=/;Max-Age=30, n$cu=1570590792583;Path=/;Max-Age=30, n$cu=1570590792583;Path=/;Max-Age=30,**

**n$cu=1570590792583;Path=/;Max-Age=30, n$cu=1570590792583;Path=/;Max-Age=30, n$cu=1570590792583;Path=/;Max-Age=30, n$cu=1570590792583;Path=/;Max-Age=30, n$cu=1570590792583;Path=/;Max-Age=30, selected\_culture\_code=zh\_CN; Expires=Thu, 08-Oct-2020 03:13:12 GMT;**

**Path=/, selected\_country\_code=CN; Expires=Thu, 08-Oct-2020 03:13:12 GMT; Path=/, JSESSIONID=8FD62FD4FA71046A76E5C47D956D7BFB.WAS\_93.WAS\_93; Path=/; HttpOnly, ROUTEID=route.WAS\_213; path=/;', 'transfer-encoding': 'chunked', 'Connection': 'keep-alive'}**

取这段字符串中**JSESSIONID=8FD62FD4FA71046A76E5C47D956D7BFB.WAS\_93.WAS\_93**

就是请求数据时用的参数.正则表达式: ‘JSESSIONID.\*?;’

步骤2:请求航班数据

请求方式:POST

请求地址:**https://www.eastarjet.com/json/dataService**

请求头:**headers = {**

**"Host": "www.eastarjet.com"**

**,"Connection": "keep-alive"**

**,"Origin": "https://www.eastarjet.com"**

**,"User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/77.0.3865.90 Safari/537.36"**

**,"Content-type": "text/plain"**

**,"Accept": "\*/\*"**

**,"Sec-Fetch-Site": "same-origin"**

**,"Referer": "https://www.eastarjet.com/newstar/PGWHC00001"**

**,"Accept-Encoding": "gzip, deflate, br"**

**,"Accept-Language": "zh-CN,zh;q=0.9"**

**,"Cookie": jsid**

**,"Content-Length": "1610"**

**}**

其中里面白色字体的jsid,就是步骤一得到的参数,直接替换进去.

注:该参数有效期至少四小时,不必每次请求数据都带新的jsid,这样效率低下.

基本上响应数据的长度小于1000个字符时,可以视为该数据是无效的,一般是请求参数错误,或者是jsid已过期..一般连续五次响应数据长度小于1000时,就可以断定jsid已过期.如果用最新的jsid请求,连续出现响应数据长度还是小于1000时,就是请求参数错误.

请求参数:

**'{"id": 11, "method": "DataService.service", "params": [{"javaClass": "com.jein.framework.connectivity.parameter.RequestParameter", "requestUniqueCode": "PGWHC00001", "requestExecuteType": "BIZ", "DBTransaction": false, "sourceName": null, "sourceExtension": null, "functionName": "DTWBA00022", "panelId": null, "methodType": null, "inParameters": {"javaClass": "java.util.List", "list": [{"javaClass": "com.jein.framework.connectivity.parameter.InParameter", "paramName": "flightSearch", "ioType": "IN", "structureType": "FIELD", "data": {"javaClass": "java.util.List", "list": [{"map": {"flightSearch": "{\"viewType\":\"B\",\"fly\_type\":\"2\",\"person1\":\"1\",\"person2\":\"0\",\"person3\":\"0\",\"residentCountry\":\"KR\",\"currency\":\"\",\"promotion\_cd\":\"\",\"flySection\":[{\"departure\_cd\":\"ICN\",\"departure\_txt\":\"\",\"arrival\_cd\":\"TPE\",\"arrival\_txt\":\"\",\"departure\_date\_cd\":\"20191002\",\"departure\_date\_txt\":\"\"}],\"recaptchaToken\":\"03AOLTBLR0-StEIwDrFmDfMK0JXuxTfkf2Gca4rLjvwEM7bPFoCfI4QeU1MDPSDik7UAvEZOadnzzkcZc1NmDo-ZoEy\_VX46Y5iebUE05GQJzsGIH-HZHbtrnvMqjNjUI-7k19Ojn5n5LcqplZT8R4-5Ys08o7I5fwRA0ri9Tg7ACVFRnb-U1QDaNDYaCuZW8oDwxt1RptMaiEplaPHUdFybsW\_km-9ksugcNZcE694-0yfS8Muuig1ivpsodqBu-44mbQesiNGxLmuKujo8GP5nO4OzA5z5AmVpEY\_q4BVsP767VAvhfb0vkEue2C7PpVRuCyncxLo01Rlu7oxPe\_YUlWWXyCLhYn2\_c6PXcIipB-PwUQnmLX1qRLzekvK5H-2B39m91f1YSKNciU4octoKNIXk7-ZLKpJGtlm0jXTjYe5GSJ2DFIWL0QWhnAXHr5uZtXpqVzcFyRjeXCRpqxgFVc8I-qkOxuyA\"}"}, "javaClass": "java.util.Map"}]}}]}, "filterParameter": {"javaClass": "java.util.Map", "map": {}}}]}'**

把这一段视为绝对的字符串,包括\转义符也不要起转义的作用,

将这段字符串中

ICN替换为出发地城市码,

TPE替换为目的城市码

20191002替换为出发日期.

步骤3:分析数据.

得到一个JSON数据后,层级关系如下

**JSON数据["result"][0]["resultData"][0]["FlightSearch"]["trips"][0]**

得到一个数组,数组长度为1则有一条航班数据,为2则两条,每一条数据都是一个字典.

key关键字:

航班号:**flightNumberText**

出发时间:**standardTimeOfDeparture**

到达时间:**standardTimeOfArrival**

正常票价:**y\_amount**

折扣票价:**d\_amount**

特价票:**e\_amount**

如果没有特价票的话,则没有e\_amount这个key

至此,数据请求完毕.

Kcc by 2019.10.9