# 计网作业: 爬虫

# 使用说明

## 环境配置方法

- 1 pip install Scrapy
- 2 scrapy startproject airport
- 3 cd airport # 进入项目airport目录
- 4 scrapy genspider wenzhou http://www.wzair.cn/lkfw/hbxx/jrdg/index.html? v=1606439135598&rxLoad=1&\_rand=1606586459057

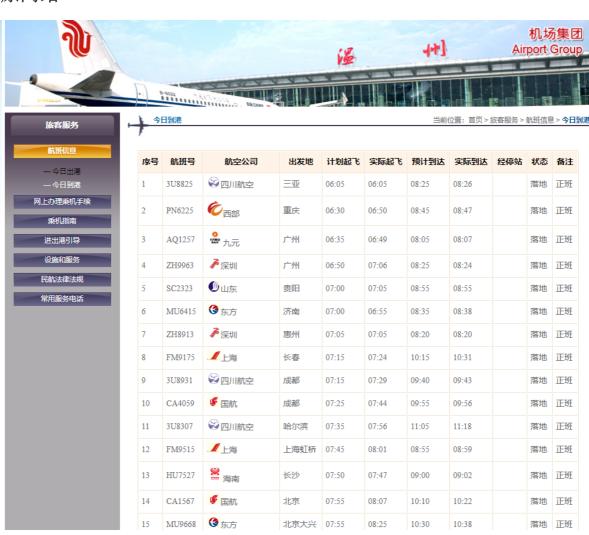
#### 代码运行说明

在最外层的airport文件夹下,输入以下命令,即可运行(每24h爬取一次数据)

1 python main.py # 使用python 3版本

# 爬取数据说明&截图

#### 原网站



```
🚺 [温州机场]2020-11-29.json
       jsonp1606647111005([
               "air_attr": "落地",
               "air company": "四川航空",
               "air_etime1": "08:25",
               "air etime2": "08:26",
               "air etime3": "08:25",
               "air from": "三亚",
               "air id": 1734569,
               "air memo": "正班",
               "air number": "3U8825",
               "air_stay": "",
  12
               "air_stime1": "06:05",
               "air stime2": "06:05",
               "air time": 1606610700000,
               "air to": "温州",
               "air_type": 0,
               "via3": "",
               "via4": ""
               "air_attr": "落地",
               "air company": "西部",
               "air etime1": "08:45"
               "air etime2": "08:47",
               "air etime3": "08:45",
               "air from": "重庆",
               "air id": 1734404,
               "air_memo": "正班",
               "air number": "PN6225",
               "air stav": ""
```

共125条

## 字段说明

以爬得的第一条数据为例:

```
1 {
2     "air_attr": "落地", // 状态
3     "air_company": "四川航空", // 航空公司
4     "air_etime1": "08:25", // 预计到达
5     "air_etime2": "08:26", // 实际到达
```

```
"air_etime3": "08:25", // 多余字段
6
 7
            "air_from": "三亚", // 出发地
            "air_id": 1734569, // 航班编号
 8
 9
            "air_memo": "正班", // 备注
            "air_number": "3U8825", // 航班号
10
           "air_stay": "", // 经停站
11
            "air_stime1": "06:05", // 计划起飞
12
13
            "air_stime2": "06:05", // 实际起飞
           "air_time": 1606610700000, // 时间
14
           "air_to": "温州", // 到达地
15
           "air_type": 0, // 类型
16
            "via3": "", //冗余字段
17
            "via4": "", //冗余字段
18
           "viac": "" //冗余字段
19
        }
 20
```