

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



旅客服务

航班信息

— 今日出港

— 今日到港

网上办理乘机手续

乘机指南

进出港引导

设施和服务

民航法律法规

常用服务电话

今日到港

当前位置：首页 > 旅客服务 > 航班信息 > 今日到港

| 序号 | 航班号 | 航空公司 | 出发地 | 计划起飞 | 实际起飞 | 预计到达 | 实际到达 | 经停站 | 状态 | 备注 |
|----|--------|------|------|-------|-------|-------|-------|-----|----|----|
| 1 | 3U8825 | 四川航空 | 三亚 | 06:05 | 06:05 | 08:25 | 08:26 | | 落地 | 正班 |
| 2 | PN6225 | 西部 | 重庆 | 06:30 | 06:50 | 08:45 | 08:47 | | 落地 | 正班 |
| 3 | AQ1257 | 九元 | 广州 | 06:35 | 06:49 | 08:05 | 08:07 | | 落地 | 正班 |
| 4 | ZH9963 | 深圳 | 广州 | 06:50 | 07:06 | 08:25 | 08:24 | | 落地 | 正班 |
| 5 | SC2323 | 山东 | 贵阳 | 07:00 | 07:05 | 08:55 | 08:55 | | 落地 | 正班 |
| 6 | MU6415 | 东方 | 济南 | 07:00 | 06:55 | 08:35 | 08:38 | | 落地 | 正班 |
| 7 | ZH8913 | 深圳 | 惠州 | 07:05 | 07:05 | 08:20 | 08:20 | | 落地 | 正班 |
| 8 | FM9175 | 上海 | 长春 | 07:15 | 07:24 | 10:15 | 10:31 | | 落地 | 正班 |
| 9 | 3U8931 | 四川航空 | 成都 | 07:15 | 07:29 | 09:40 | 09:43 | | 落地 | 正班 |
| 10 | CA4059 | 国航 | 成都 | 07:25 | 07:44 | 09:55 | 09:56 | | 落地 | 正班 |
| 11 | 3U8307 | 四川航空 | 哈尔滨 | 07:35 | 07:56 | 11:05 | 11:18 | | 落地 | 正班 |
| 12 | FM9515 | 上海 | 上海虹桥 | 07:45 | 08:01 | 08:55 | 08:59 | | 落地 | 正班 |
| 13 | HU7527 | 海南 | 长沙 | 07:50 | 07:47 | 09:00 | 09:02 | | 落地 | 正班 |
| 14 | CA1567 | 国航 | 北京 | 07:55 | 08:07 | 10:10 | 10:22 | | 落地 | 正班 |
| 15 | MU9668 | 东方 | 北京大兴 | 07:55 | 08:25 | 10:30 | 10:38 | | 落地 | 正班 |

截图

```
{ [温州机场]2020-11-29.json
1  jsonp1606647111005([
2      {
3          "air_attr": "落地",
4          "air_company": "四川航空",
5          "air_etime1": "08:25",
6          "air_etime2": "08:26",
7          "air_etime3": "08:25",
8          "air_from": "三亚",
9          "air_id": 1734569,
10         "air_memo": "正班",
11         "air_number": "3U8825",
12         "air_stay": "",
13         "air_stime1": "06:05",
14         "air_stime2": "06:05",
15         "air_time": 1606610700000,
16         "air_to": "温州",
17         "air_type": 0,
18         "via3": "",
19         "via4": "",
20         "viac": ""
21     },
22     {
23         "air_attr": "落地",
24         "air_company": "西部",
25         "air_etime1": "08:45",
26         "air_etime2": "08:47",
27         "air_etime3": "08:45",
28         "air_from": "重庆",
29         "air_id": 1734404,
30         "air_memo": "正班",
31         "air_number": "PN6225",
32         "air_stay": "",
```

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
```

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