

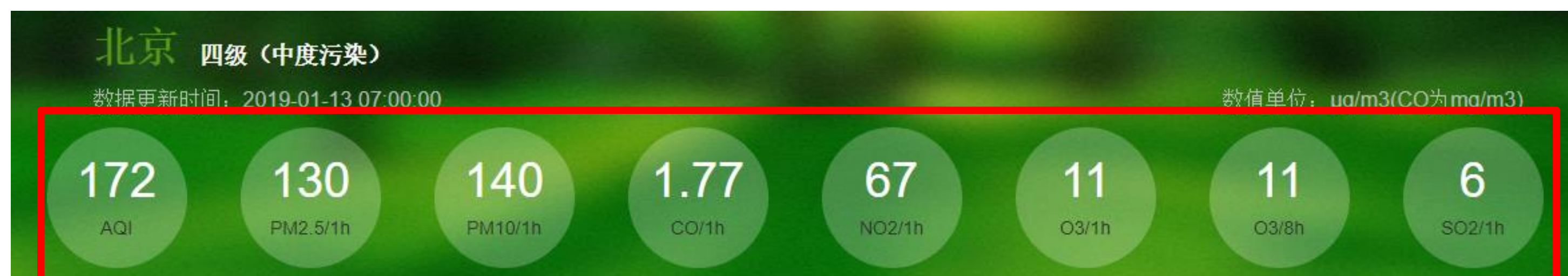
第三课 网络数据的获取与表示

第11节 实战：获取国内城市空气质量指数数据

主 讲：Robin

获取国内城市空气质量指数数据

- URL: <http://www.pm25.in/>
- 任务：爬取国内所有城市的AQI数据，并保存到本地
- 网站结构：
 - <http://www.pm25.in/> 包含所有城市的名称和链接
 - 二级链接包含具体城市的各项AQI数据，如
<http://www.pm25.in/beijing> 包含



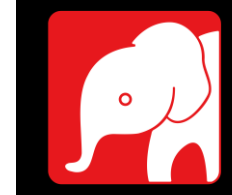
获取国内城市空气质量指数数据

- 步骤:

1. 创建项目 scrapy startproject **aqi_proj**
2. 定义Item, **CityAqiItem()**

```
class CityAqiItem(scrapy.Item):
    city_name = scrapy.Field() # 城市名称
    city_link = scrapy.Field() # 城市链接
    aqi = scrapy.Field() # AQI
    pm25 = scrapy.Field() # PM2.5
    pm10 = scrapy.Field() # PM10
    co = scrapy.Field() # CO
    no2 = scrapy.Field() # NO2
    o3_1h = scrapy.Field() # O3/1h
    o3_8h = scrapy.Field() # O3/8h
    so2 = scrapy.Field() # SO2
```

3. 编写Spider, 解析**首页URL**及**二级链接**
4. 编写Pipeline, 结果保存到CSV文件



只为遇见明天更优秀的你！