

第三课 网络数据的获取与表示

第八节 Scrapy使用步骤 (2)

主 讲: Robin

Scrapy使用步骤

- 安装: `conda install -c conda-forge scrapy`
- 1. 创建工程
- 2. 定义Item, 构造爬取的对象 (可选)
- 3. 编写Spider, 爬虫主体
- 4. 编写配置和Pipeline, 用于处理爬取的结果 (可选)
- 5. 执行爬虫

Scrapy使用步骤

3. 编写Spider

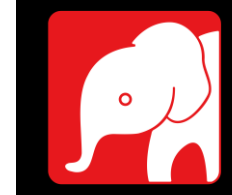
- cd tutorial
- scrapy genspider aqi_spider <http://www.pm25.in/>

```
class AqiSpider(scrapy.Spider):
    name = 'aqi_spider'
    allowed_domains = ['http://www.pm25.in/']
    start_urls = ['http://www.pm25.in/']

    def parse(self, response):
        print('=====')
        print('type:', type(response))
        print('status:', response.status)
        print('body:')
        print(response.body)
        print('=====')
```

5. 运行Spider

- scrapy crawl aqi_spider



只为遇见明天更优秀的你！