

第三课 网络数据的获取与表示

第十节 Scrapy使用步骤 (4)

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

4. 编写Pipeline

- 目的：处理解析的Item，如结果保存为CSV文件
- 在settings.py中配置

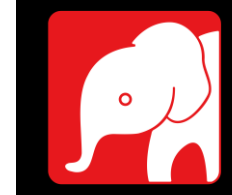
```
# Configure item pipelines
# See https://doc.scrapy.org/en/latest/topics/item-pipeline.html
ITEM_PIPELINES = {
    'tutorial.pipelines.CityItem': 300,
}
```

- 添加open_spider()函数，close_spider()函数

```
class CityItemPipeline(object):
    def open_spider(self, spider):
        self.file = open('cities.csv', 'wb')
        self.exporter = CsvItemExporter(self.file)
        self.exporter.start_exporting()

    def close_spider(self, spider):
        self.exporter.finish_exporting()
        self.file.close()

    def process_item(self, item, spider):
        self.exporter.export_item(item)
        return item
```



只为遇见明天更优秀的你！