

极客学院
jikexueyuan.com

定向爬虫 Scrapy与Redis入门

■ 定向爬虫：Scrapy与Redis入门 — 课程概要

- Redis的介绍与安装
- Scrapy配置Redis
- 实战——小说爬虫v0.2

Redis的介绍与安装

Redis的介绍与安装

- Redis介绍
- 安装scrapy-redis模块
- 安装与运行Redis

Redis的介绍与安装 — Redis的介绍

Redis 是一个高性能的key-value数据库。它将数据保存在内存中，因此可以实现非常快的存取速度。



Redis is an open source, BSD licensed, advanced key-value cache and store. It is often referred to as a data structure server since keys can contain strings, hashes, lists, sets, sorted sets, bitmaps and hyperloglogs.

——Redis Home Page

Redis的介绍与安装 — 安装scrapy-redis模块

- `pip install scrapy-redis`
- `easy_install scrapy-redis`

Redis的介绍与安装 — 安装与运行Redis

安装运行Redis

- 打开网址: <http://redis.io/download>
- 运行Redis: `redis-server redis.conf`
- 清空缓存: `redis-cli flushdb`

Scrapy配置Redis

Scrapy配置Redis

- settings.py配置Redis
- Spider 调用Redis

Scrapy配置Redis— **settings.py**配置Redis

```
SCHEDULER = "scrapy_redis.scheduler.Scheduler"
SCHEDULER_PERSIST = True
SCHEDULER_QUEUE_CLASS = scrapy_redis.queue.SpiderPriorityQueue'
REDIS_URL = None
REDIS_HOST = '127.0.0.1'
REDIS_PORT = 6379
```

Scrapy配置Redis— Spider调用Redis

```
from scrapy_redis.spiders import RedisSpider  
class xxx(RedisSpider):  
.....
```

实战——小说爬虫v0.2

实战——小说爬虫

目标网站：盗墓笔记小说网站

目标网址：<http://www.daomubiji.com/>

目标内容：

盗墓笔记小说的信息与内容，具体内容包括：

- 书标题
- 章数
- 章标题
- 每一章正文

输出结果保存到MongoDB中。

定向爬虫：Scrapy与Redis 入门

本套课程中我们学习了Redis在定向爬虫中的应用，你应当掌握以下知识：

- 安装和使用Redis
- Scrapy调用Redis
- 将爬取结果保存到MongoDB中

掌握Redis是实现定向爬虫的方法之一。如果想继续提高，你可以继续在极客学院学习《定向爬虫入门》课程。

极客学院

jikexueyuan.com

中国最大的IT职业在线教育平台

