# 基于 Scrapy 框架影视信息采集与分析

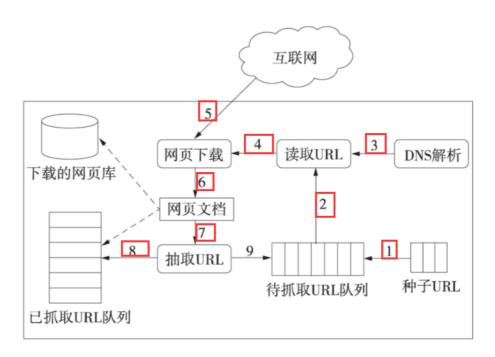
## 项目介绍

为了充分利用网上大数据资源,让用户能够方便利用影视信息,采用基于 Scrapy 框架的爬虫技术,开发

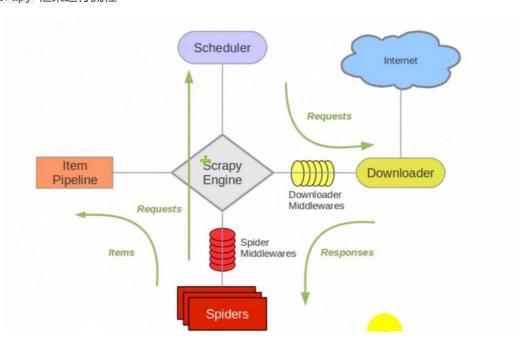
了检索电影信息的搜索引擎。对豆瓣网站的影视信息进行爬取,以方便用户准确获取最新的电影信息。

### 项目流程图

• 通用爬虫框架流程



• Scrapy 框架运行流程



# 项目代码

以"豆瓣电影"为爬取目标,爬取网站中的影视信息。主要包括网站排名 "Top250"和喜剧、动作类电影的电影名称、电影评分、电影导演,电影上映时间以及电影评语。

#### 创建工程

scrapy startproject DouBan

## 创建爬虫程序

cd DouBan/
scrapy genspider douban 'douban.com'

• 自动创建目录及文件