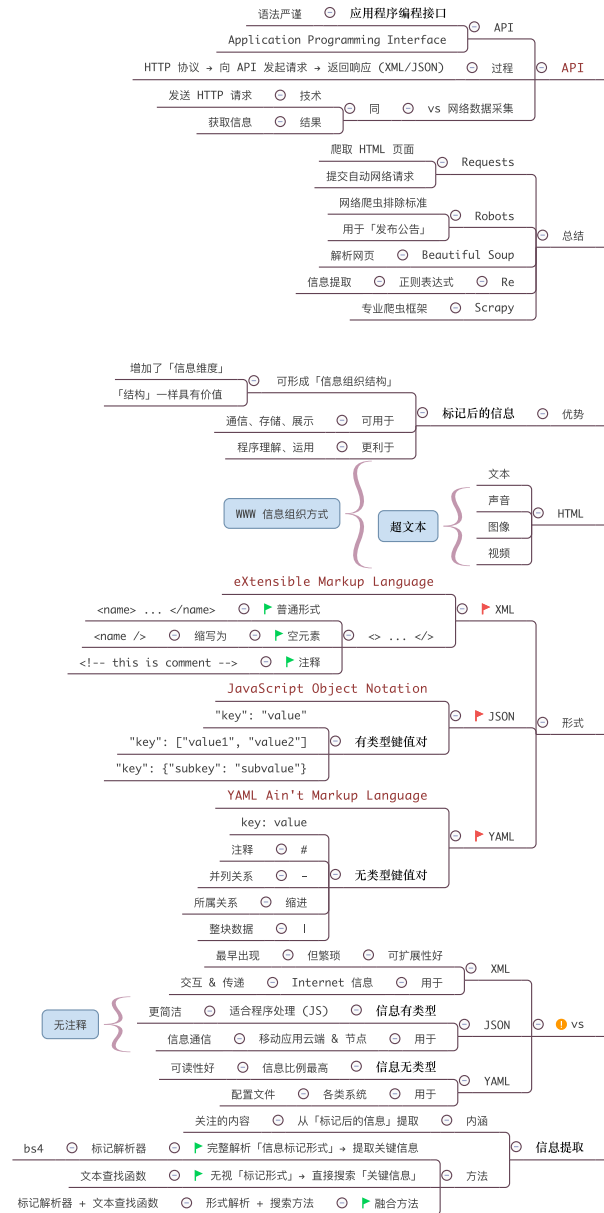


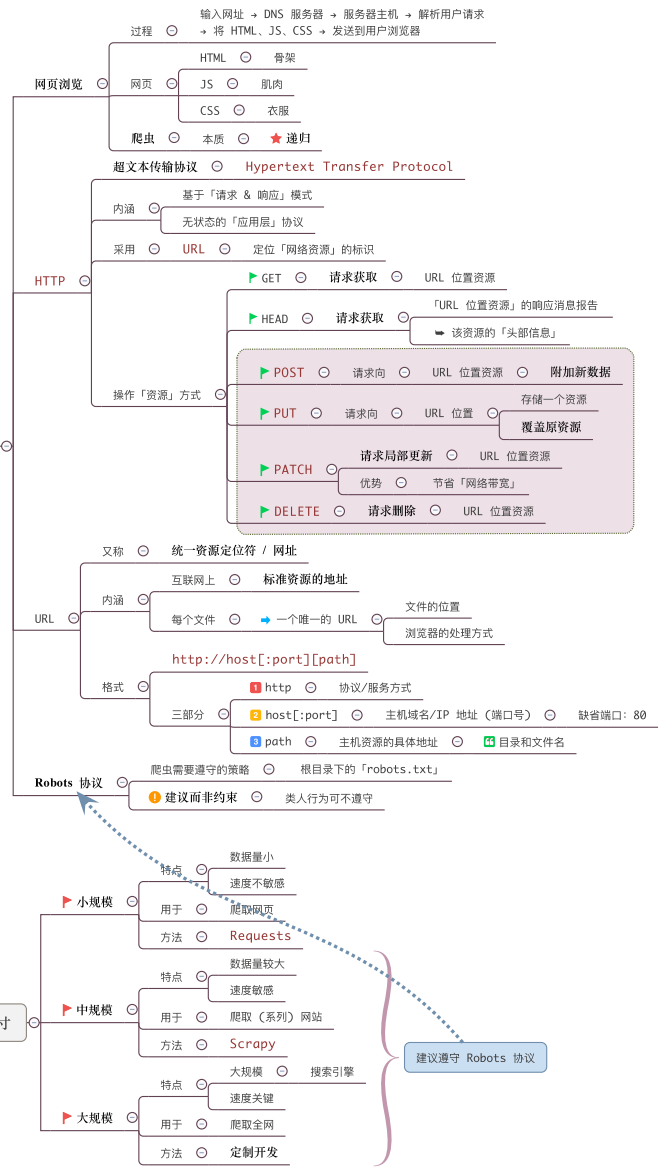
ZOE's MindMap

1-4 | v2.0 | 20180426



延伸知识

背景知识



公众号: 数林觅风

ZOE's MindMap

2-4 | v2.0 | 20180426

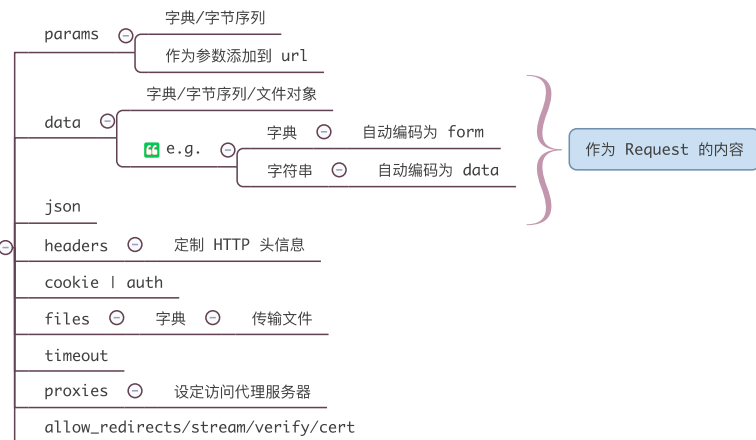
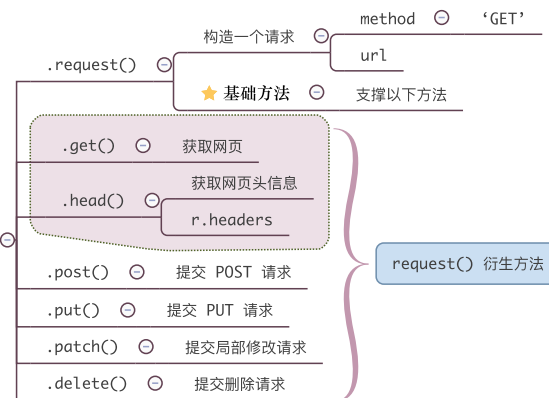
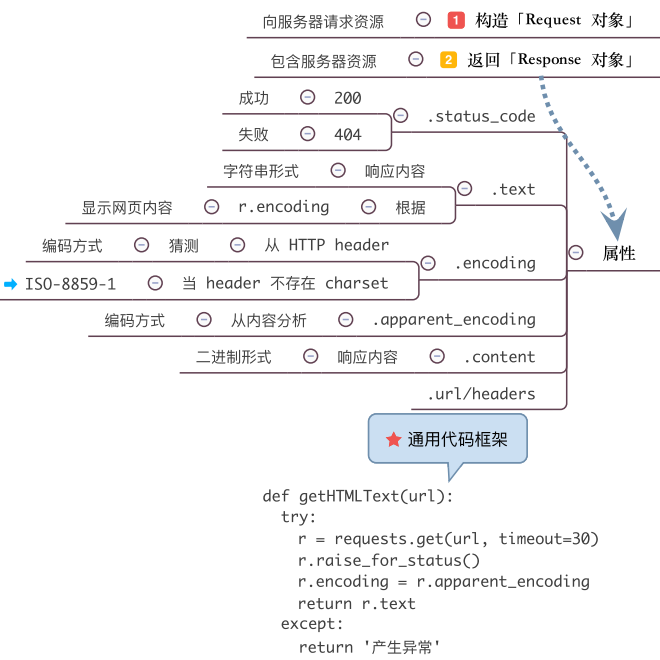
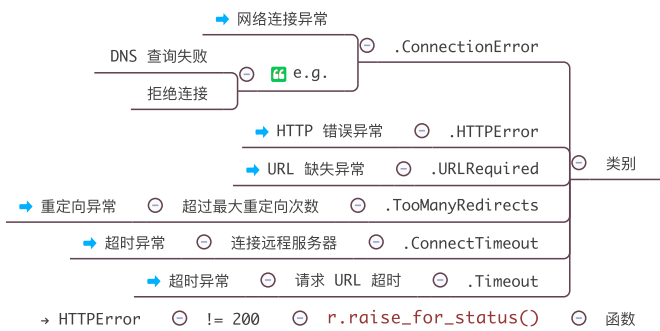
公众号: 数林觅风
<https://woaielf.github.io/>
2 Requests 库

异常

步骤

主要方法

控制访问参数



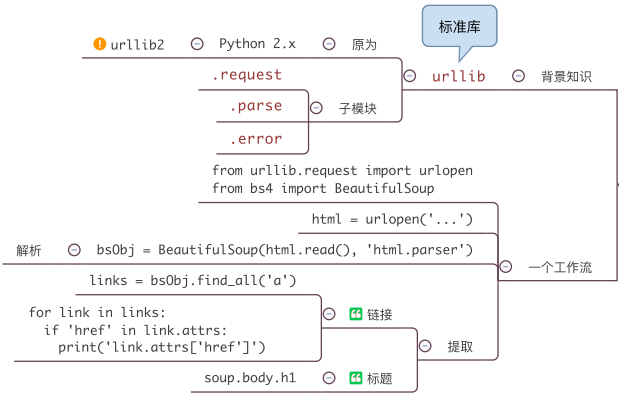
公众号: 数林觅风

ZOE's MindMap

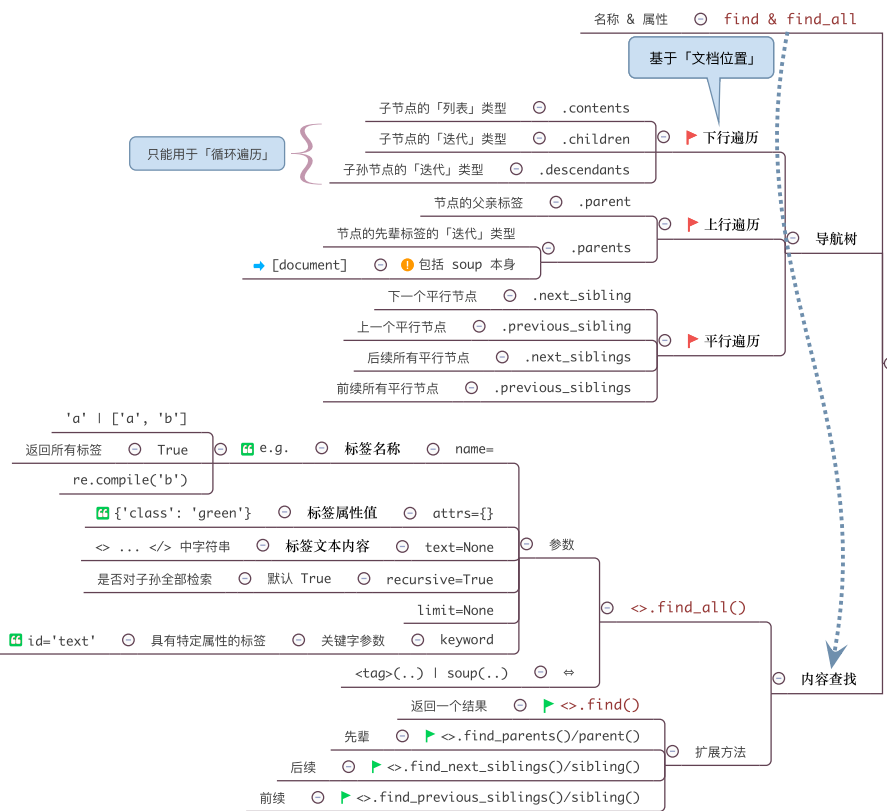
3-4 | v2.0 | 20180426

公众号: 数林觅风
<https://woaielf.github.io/>
3 BeautifulSoup & urllib

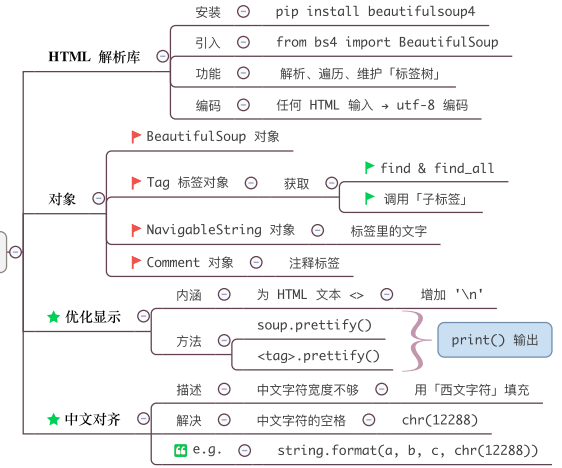
urllib



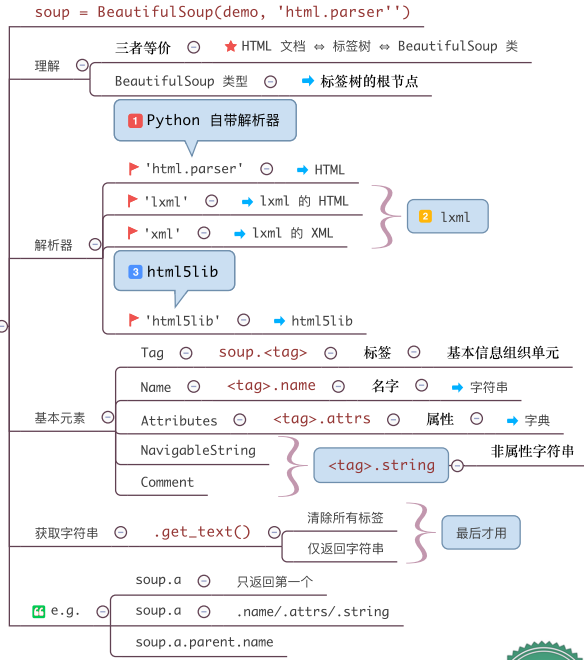
查找方式



BeautifulSoup



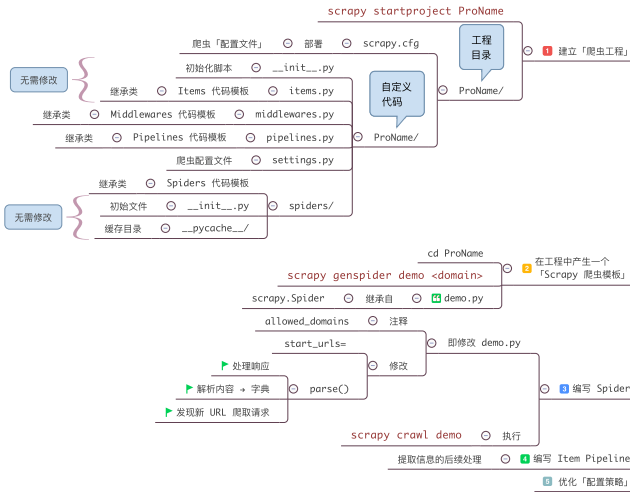
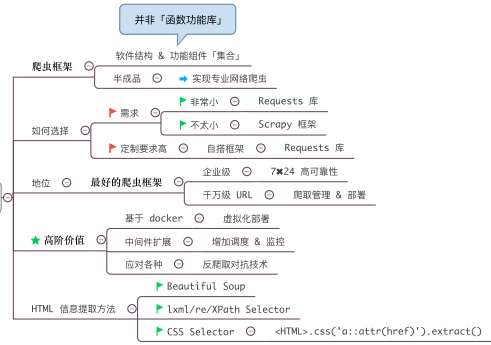
基本方法



Requests	Scrapy	同! !
页面级 功能库 并发性不足 性能差 重点: 页面下载 定制灵活	网站级 框架 并发性好 性能高 重点: 爬虫结构 一般定制 OK, 深度定制困难	页面请求 & 爬取 同为两条重要技术路线 requests-bs4-re scrapy 处理 js 提交表单 应对验证码 PhantomJS

Requests vs Scrapy

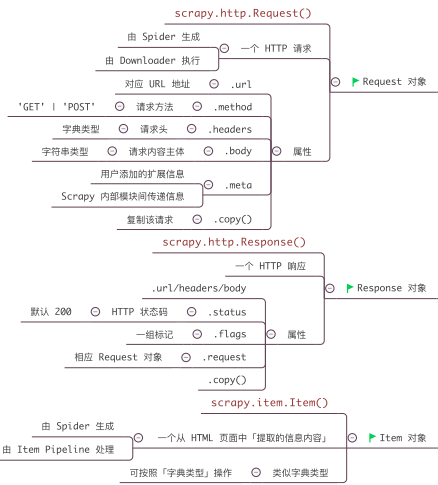
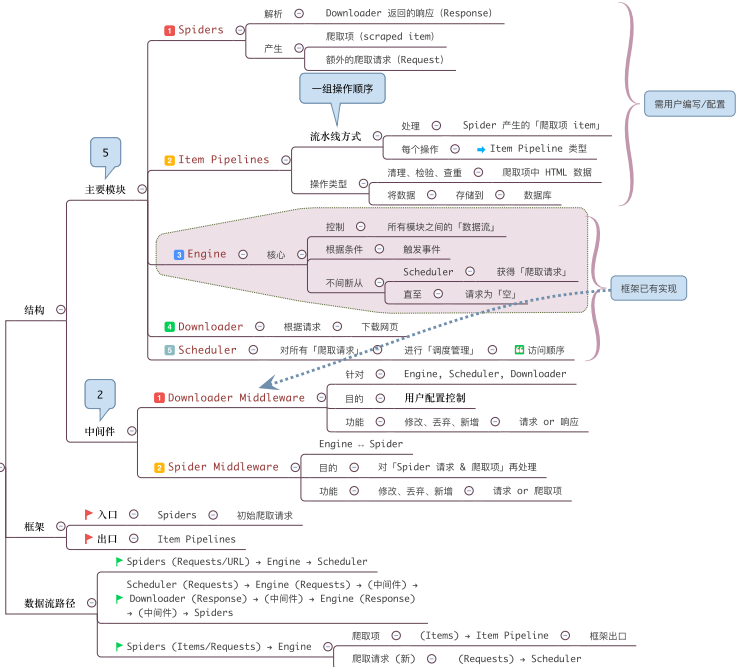
简介



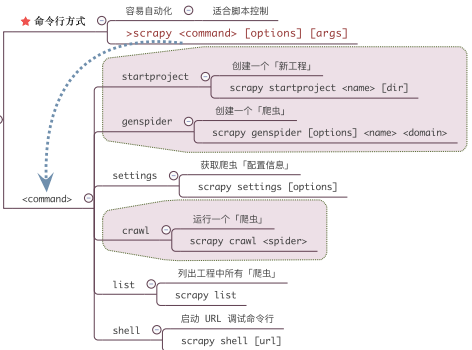
使用步骤

公众号: 数林梵风
<https://woaieif.github.io/>
4 Scrapy 爬虫框架

“5+2” 结构



常用命令



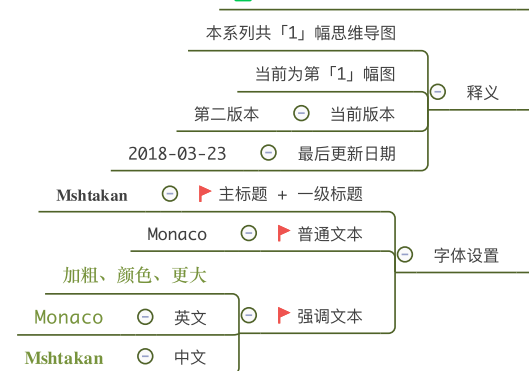
ZOE's MindMap

1-1 | v2.1 | 20180402



编号-系列总数 | 版本号 | 更新日期

1-1 | v2.0 | 20180323

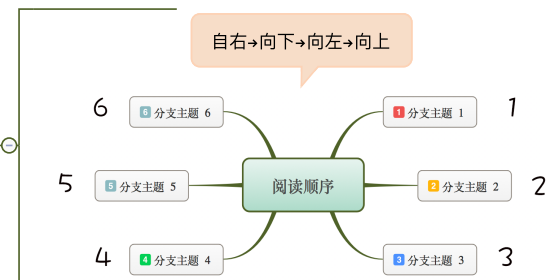


4 关于作者

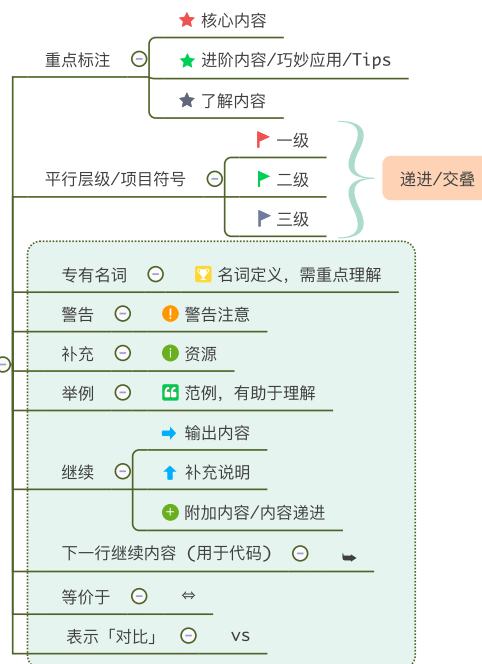
公众号：数林觅风
<https://woaielf.github.io/>
❤️ ZOE 思维导图规则

1 阅读顺序

顺时针方向！



2 ★ 图标含义



图例

- 1 编号1
- 2 编号2
- 3 编号3
- 4 编号4
- 一级
- 二级
- 三级
- 核心理解
- 进阶内容
- 简单了解
- 输出内容
- 补充前面
- 注意
- 附加/递进
- 资源
- 举例说明
- 爱心
- 名词定义



公众号：数林觅风



获取更多 ZOE 思维导图

❤️ <https://woaielf.github.io/>

👤 数林觅风 | zoemindmap

📺 @数林觅风

📖 ZOE 酱



思维导图 | 数据科学 | 编程语言 | 读书笔记 | 精进思考

