p193 find（）

p194 7 . css 选择器 嵌套选择

p195 获取属性 获取文本

p196 4.3 使用pyquery 1 . 准备工作 2. 初始化 字符串初始化

p197 URL 初始化 文件初始化 3. 基本css 选择器

p198 4. 查找节点 子节点

p199 父节点

p201 兄弟节点 5. 遍历

p202 6. 获取信息 获取属性

p203 获取文本

p205 7. 节点操作 addClass 和removeClass

p206 attr 、text 和html remove()

p207 8. 伪类选择器

p209 第五章 数据存储 5.1 文件存储 5.1.1 TXT 文本存储 1. 本节目标 2 . 基本实例

p210 3. 打开方式

p211 4. 简化写法 5.1.2 JSON 文件存储 1 . 对象和数组

p212 2. 读取JSON

p214 3. 输出JSON

p215 5. 1.3 csv 文件存储

p216 1. 写入

p218 2 . 读取

p219 5.2关系型数据库存储 5.2.1MySQL 的存储 1. 准备工作 2 . 连接数据库

p220 3 . 创建表 4. 插入数据

p223 5. 更新数据

p224 6 . 删除数据 7 . 查询数据

p225 5.3 非关系型数据库存储

p226 5.3.1 MongoDB 存储 1 . 准备工作 2. 连接MongoDB 3. 指定数据库 4 . 指定集合

p227 5. 插入数据

p228 6. 查询

p230 7. 计数 8 排序 9 . 偏移

p231 10. 更新

p232 1 1 . 删除

p233 12. 其他操作 5.3. 2 Redis 存储 1. 准备工作 2. Redis 和Strict Redis 3. 连接Redis

p234 4. 键操作

p235 5. 字符串操作

p237 6 . 列表操作

p238 7 集合操作

p239 8. 有序集合操作

p241 9 . 散列操作

p242 10. RedisDump redis-dump

p243 redis-load

p244第6章Ajax数据爬取 6.1 什么是Ajax

245 1. 实例引入 2. 基本原理 ·发送请求

246 解析内容 渲染网页 6.2 Ajax 分析方法

247 1. 查看请求

249 2. 过滤请求

250 6.3 Ajax 结果提取

251 1. 分析请求 2. 分析响应

252 3. 实战演练

254 6.4 分析Ajax 爬取今日头条街拍美图 1 准备工作

255 2. 抓取分析

258 3 实战演练

261 第7章 动态渲染页面爬取 7.1 Selenium 的使用 1 . 准备工作 2 . 基本使用

263 3 . 声明浏览器对象 4. 访问页面 5 . 查找节点 单个节点

265 多个节点

266 6 . 节点交互

267 7. 动作链 8. 执行JavaScript

268 9. 获取节点信息 获取属性 获取文本值

269 获取id、位置、标签名和大小 10. 切换Frame

270 11 . 延时等待 隐式等待 显式等待

272 12 . 前进和后退 13. Cookies

273 14. 选项卡管理 15. 异常处理

274 7.2 Splash 的使用 1 功能介绍

275 2. 准备工作 3 . 实例引入

276 4 . Splash Lua 脚本 入口及返回值

277 异步处理

278 5 . Splash 对象属性 args

279 js\_enabled resource timeout

280 images\_ enabled plugins \_enabled scroll\_position

281 6. Splash 对象的方法 go()

282 wait()

283 evaljs() runjs() autoload()

284 call\_later()

285 http\_get() http post()

286 set\_content()

287 html() png() jpeg()

288 har() url() get\_cookies()

289 add\_cookie() clear\_ cookies() get\_viewport\_size()

290 set\_ viewport size() set\_viewport\_full() set\_user\_agent()

291 set\_custom\_headers() select()

292 select\_all()

293 mouse\_click()

294 7. Splash API 调用 render.html render.pug

295 render.jpeg render.bar

296 render.json execute

298 7.3 Splash 负载均衡配置 1. 配置Splash 服务 2 . 配置负载均衡

299 3. 配置认证

300 4 . 测试

301 7.4 使用Selenium 爬取淘宝商品 1. 本节目标 2. 准备工作 3 . 接口分析

302 4. 页面分析

303 5. 获取商品列表

305 6 . 解析商品列表

306 7 . 保存到MongoDB

307 8 . 遍历每页 9. 运行

308 10. Chrome Headless 模式 11对接Firefox12对接PhantomJS

310 第8章 验证码的识别