

web scraper正确操作流程

软件介绍

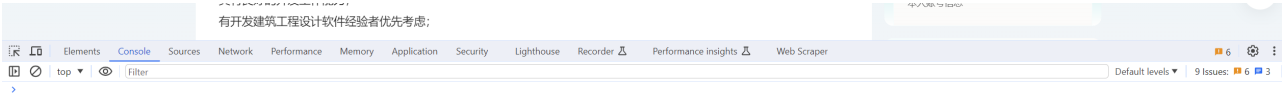
web scraper是谷歌浏览器的一个爬虫插件，能够让完全不懂爬虫的人无障碍实现爬虫

下载和安装

自行到谷歌浏览器的拓展插件处下载

使用教程

1. 快捷键F12，打开网页源码界面

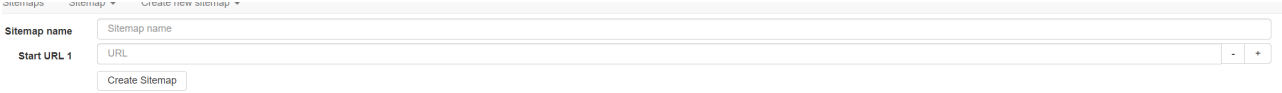


2. 选择最右侧的选项卡

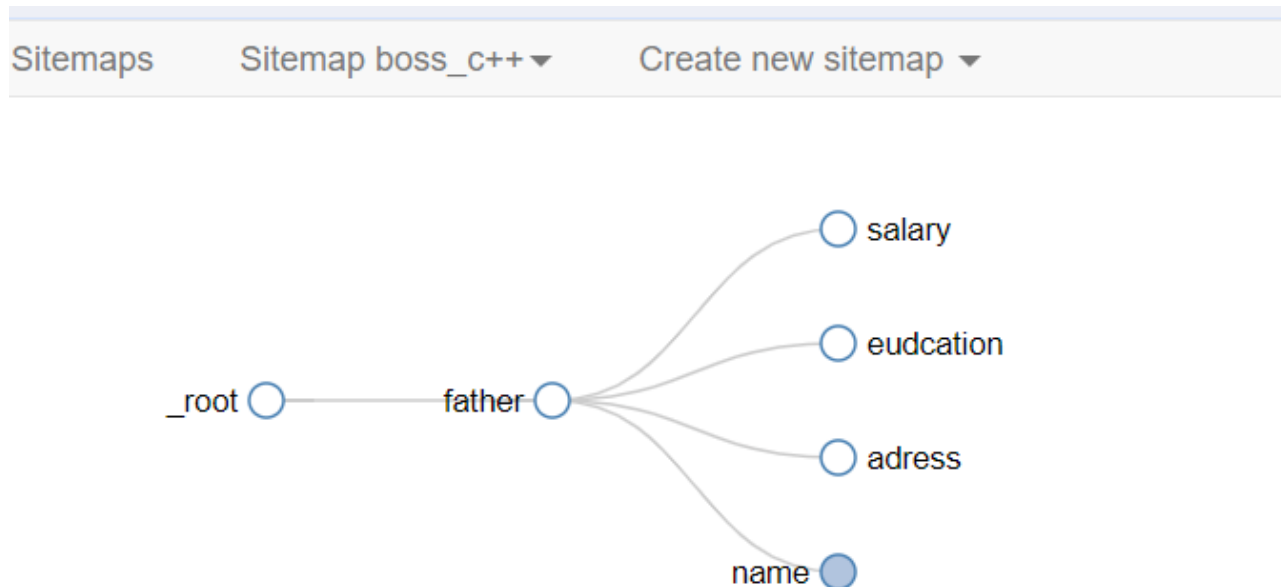
3. 选择创建或者导入抓取任务



4. 创建抓取任务要制定抓取的网页（URL）



5. 接下来就是通过不同层级的选择器来选择抓取内容



抓取的内容满足父子关系，以上图为例：

father抓取的是每个职位的简介卡片，其孩子的内容就是选项卡内的某些内容，包括职位名称、工作地点、学历要求、工资水平

6. 对于每个选择器可以通过点击select，然后鼠标点击要选择的区域，选择两个及以上的不同区域，会自动选择当前页面的所有相同类型的区域

The screenshot shows a web scraper interface with a list of job listings. The first listing is for a C++ position in Beijing, Shijingshan District, Apple Garden, with a salary of 20-28K, 3-5 years experience, and a degree. The second listing is for a C++ position in Beijing, Daxing District, Yizhuang, with a salary of 15-25K, 1-3 years experience, and a degree. The third listing is for a C++ position in Beijing, Haidian District, Gaojunmiao, with a salary of 25-30K, 5-10 years experience, and a degree. A tooltip shows the URL for the first listing: `/3222a9a070f4e0081nZ73N-5GFJQ.html?ch.272&securityId=k8FlfwFuQA0pf-s1CQG2Kzmi8NRmXz94Uc5HhrCRMcIBJV-nJWZY3Eq0MHlgeW-cT3sQXd3fPj0lvp6zmBy74C7ABf0MmIoG&sessionId=" ka="search_list_jname_272_blank"s="job-card-left -ws-data-preview-element">`. Below the listings, there are buttons for 'P', 'C', 'S', and 'Done selecting'. At the bottom, there is a 'Selector configuration' panel with fields for 'Selector Id', 'Text', and 'Select'. The 'Select' field is set to 'Element preview'.

7. father选择element类型，勾选multiple，选择父选择器为根选择器

The screenshot shows the 'Selector configuration' panel. The 'Id' field is set to 'father'. The 'Text' field is set to 'Element'. The 'Select' field is set to 'Element preview'. The 'Multiple' checkbox is checked. The 'Root' field is set to 'father'. The 'Name' field is set to 'name'. The 'Save selector' button is highlighted.

element类型是一个大盒子，并不是具体的数据内容，需要靠子选择器来选择盒子里面的内容

8. 用同样的方式，设置好子选择器的内容。注意，子选择器不需要勾选multiple

salary	span.salary	SelectorText	no	father	Element preview	Data preview	Edit	Delete
education	li:nth-of-type(2)	SelectorText	no	father	Element preview	Data preview	Edit	Delete
address	span.job-area	SelectorText	no	father	Element preview	Data preview	Edit	Delete
name	_parent_	SelectorLink	no	father	Element preview	Data preview	Edit	Delete

9. 在第二个选项卡中可以执行爬虫，导出爬虫数据

🏠🔍

ElementsConsoleSourcesNetworkPerformanceMemory

Sitemaps

Sitemap boss_c++▼

Create new sitemap ▼

_root / father

	ID
☰	salary
☰	eudcation
☰	adress
☰	name

Selectors

Selector graph

Edit metadata

Scrape

Browse

Export Sitemap

Export data

	Selector
	span.salary
	li:nth-of-type(2)
	span.job-area
	parent

Add new selector

10. 最终效果

25-35K 13薪	本科	北京·朝阳区·朝外	跨平台 C++ 软件研发工程师北京·朝阳区·朝外25-35K 13薪104https://www	<p>岗位职责:</p> <p>使用C/C++ 语言开发音频传输、音频处理研发工作;</p> <p>负责终端 SDK 的接口定义与封装实现, 以及多平台支持和适配;</p> <p>职位要求:</p> <p>1. 统招全日制本科及以上学历, 计算机软件、通讯相关专业;</p> <p>2. 良好的数据结构和算法能力, 精通 C/C++ 开发; 熟练使用QT或JUICE等UI软件编程工具; 对OpenGL有了解;</p> <p>3. 精通多线程和网络编程, 有音视频或流媒体项目开发经验, 涉及UDP组播传输, RTP、RTCP传输协议、网络协议、信号传输、音视频监控类项目经验;</p> <p>4. 具备Electron/Cocos2d-x/Unity3D等跨平台框架开发经验;</p> <p>5. 具备强烈的进取心和责任感, 极强的学习能力及良好的团队合作精神;</p> <p>6. 相关工作经验10年左右;</p>
20-40K 15薪	本科	北京·大兴区·亦庄	C++ 推荐引擎工程师北京·大兴区·亦庄20-40K 15薪3-5年本科https://www	<p>主要职责:</p> <p>1. 参与推荐引擎、模型预测、向量检索、调度系统等基础系统和平台的设计、研发及优化工作;</p> <p>2. 研究分布式信息检索的服务架构, 分析和优化相关性算法、策略, 构建高可用性、高可扩展性的分布式检索系统;</p> <p>3. 深入理解业务, 助力业务快速发展的同时, 抽象和设计平台化解决方案, 完善整个基础技术体系, 提升迭代效率和创造价值。</p> <p>岗位要求:</p> <p>1. 算法和数据结构基础扎实, 熟练掌握Java、C/C++语言;</p>