



Puppeteer的入门教程和实践



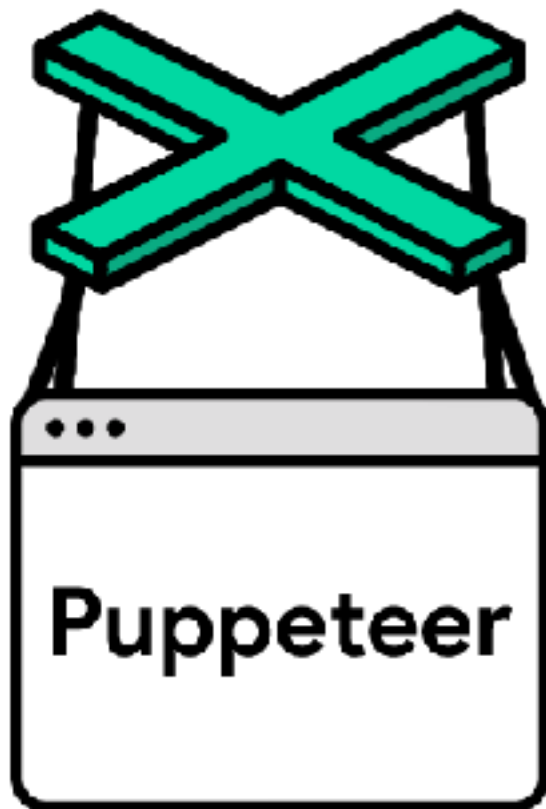
任乃千

[+ 关注](#)

2017.10.22 18:04* 字数 1187 阅读 4863 评论 2 喜欢 11

出现的背景

Chrome59(linux、macos)、Chrome60(windows)之后，Chrome自带headless(无界面)模式很方便做自动化测试或者爬虫。但是如何和headless模式的Chrome交互则是一个问题。通过启动Chrome时的命令行参数仅能实现简易的启动时初始化操作。Selenium、Webdriver等是一种解决方案，但是往往依赖众多，不够扁平。



Puppeteer.png

Puppeteer是谷歌官方出品的一个通过DevTools协议控制headless Chrome的Node库。可以通过Puppeteer的提供的api直接控制Chrome模拟大部分用户操作来进行UI Test或者作为爬虫访问页面来收集数据。



Puppeteer本身依赖6.4以上的Node，但是为了异步超级好用的`async/await`，推荐使用7.6版本以上的Node。另外headless Chrome本身对服务器依赖的库的版本要求比较高，centos服务器依赖偏稳定，v6很难使用headless Chrome，提升依赖版本可能出现各种服务器问题（包括且不限于无法使用ssh），最好使用高版本服务器。

Puppeteer因为是一个npm的包，所以安装很简单：

```
npm i puppeteer
```

或者

```
yarn add puppeteer
```

Puppeteer安装时自带一个最新版本的Chromium，可以通过设置环境变量或者npm config中的PUPPETEER_SKIP_CHROMIUM_DOWNLOAD跳过下载。如果不下载的话，启动时可以通过puppeteer.launch([options])配置项中的executablePath指定Chromium的位置。

使用和例子

Puppeteer类似其他框架，通过操作Browser实例来操作浏览器作出相应的反应。

```
const puppeteer = require('puppeteer');

(async () => {
  const browser = await puppeteer.launch();
  const page = await browser.newPage();
  await page.goto('http://rennaiqian.com');
  await page.screenshot({path: 'example.png'});
  await page.pdf({path: 'example.pdf', format: 'A4'});
  await browser.close();
})();
```

上述代码通过puppeteer的launch方法生成了一个browser的实例，对应于浏览器，launch方法可以传入配置项，比较有用的是在本地调试时传入{headless:false}可以关闭headless模式。



browser.newPage方法可以打开一个新选项卡并返回选项卡的实例page，通过page上的各种方法可以对页面进行常用操作。上述代码就进行了截屏和打印pdf的操作。

一个很强大的方法是page.evaluate(pageFunction, ...args)，可以向页面注入我们的函数，这样就有了无限可能

```
const puppeteer = require('puppeteer');

(async () => {
  const browser = await puppeteer.launch();
  const page = await browser.newPage();
  await page.goto('http://rennaiqian.com');

  // Get the "viewport" of the page, as reported by the page.
  const dimensions = await page.evaluate(() => {
    return {
      width: document.documentElement.clientWidth,
      height: document.documentElement.clientHeight,
      deviceScaleFactor: window.devicePixelRatio
    };
  });

  console.log('Dimensions:', dimensions);
  await browser.close();
})();
```

需要注意的是evaluate方法中是无法直接使用外部的变量的，需要作为参数传入，想要获得执行的结果也需要return出来。因为是一个开源一个多月的项目，现在项目很活跃，所以使用时自行查找api才能保证参数、使用方法不会错。

调试技巧

1. 关掉无界面模式，有时查看浏览器显示的内容是很有用的。使用以下命令可以启动完整版浏览器：

```
const browser = await puppeteer.launch({headless: false})
```

2. 减慢速度，slowMo选项以指定的毫秒减慢Puppeteer的操作。这是另一个看到发生了什么的方法：



```
slowMo:250  
});
```

3.捕获console的输出,通过监听console事件。在page.evaluate里调试代码时这也很方便：

```
page.on('console', msg => console.log('PAGE LOG:', ...msg.args));  
await page.evaluate(() => console.log(`url is ${location.href}`));
```

4.启动详细日志记录，所有公共API调用和内部协议流量都将通过puppeteer命名空间下的debug模块进行记录

```
# Basic verbose logging  
env DEBUG="puppeteer:*" node script.js  
  
# Debug output can be enabled/disabled by namespace  
env DEBUG="puppeteer:*,-puppeteer:protocol" node script.js # everything BUT protocol messages  
env DEBUG="puppeteer:session" node script.js # protocol session messages (protocol messages)  
env DEBUG="puppeteer:mouse,puppeteer:keyboard" node script.js # only Mouse and Keyboard API  
  
# Protocol traffic can be rather noisy. This example filters out all Network domain messages  
env DEBUG="puppeteer:*" env DEBUG_COLORS=true node script.js 2>&1 | grep -v '"Network'"
```

爬虫实践

很多网页通过user-agent来判断设备，可以通过page.emulate(options)来进行模拟。options有两个配置项，一个为用户Agent，另一个为viewport可以设置宽度(width)、高度(height)、屏幕缩放(deviceScaleFactor)、是否是移动端(isMobile)、有无touch事件(hasTouch)。

```
const puppeteer = require('puppeteer');  
const devices = require('puppeteer/DeviceDescriptors');  
const iPhone = devices['iPhone 6'];  
  
puppeteer.launch().then(async browser => {  
  const page = await browser.newPage();  
  await page.emulate(iPhone);  
  await page.goto('https://www.example.com');  
  // other actions...  
  await browser.close();  
});
```



很多网页需要登录，有两种解决方案：

1. 让puppeteer去输入账号密码

常用方法：点击可以使用`page.click(selector[, options])`方法，也可以选择聚焦`page.focus(selector)`。

输入可以使用`page.type(selector, text[, options])`输入指定的字符串，还可以在options中设置delay缓慢输入更像真人一些。也可以使用`keyboard.down(key[, options])`来一个字符一个字符的输入。

2. 如果是通过cookie判断登录状态的可以通过`page.setCookie(...cookies)`，想要维持cookie可以定时访问。

Tip：有些网站需要扫码，但是相同域名的其他网页却有登录，就可以尝试去可以登录的网页登录完利用cookie访问跳过扫码。

简单例子

```
const puppeteer = require('puppeteer');

(async () => {
  const browser = await puppeteer.launch({headless: false});
  const page = await browser.newPage();
  await page.goto('https://baidu.com');
  await page.type('#kw', 'puppeteer', {delay: 100});
  page.click('#su')
  await page.waitFor(1000);
  const targetLink = await page.evaluate(() => {
    return [...document.querySelectorAll('.result a')].filter(item => {
      return item.innerText && item.innerText.includes('Puppeteer的入门和实践')
    }).toString()
  });
  await page.goto(targetLink);
  await page.waitFor(1000);
  browser.close();
})();
```



records.gif

求赞，另外欢迎访问[我的博客](#)

小礼物走一走，来简书关注我

赞赏支持

web前端

举报文章 © 著作权归作者所有



任乃千 ♂

写了 25166 字，被 132 人关注，获得了 215 个喜欢

+ 关注

小米科技 前端工程师 掘金专栏：<https://juejin.im/user/584a3d8fb123db00660a77aa> 个人网站：<http://www...>





下载简书 App ▶
随时随地发现和创作内容



被以下专题收入，发现更多相似内容



前端学习之路



Web前端之路



让前端飞



程序员



非前端的奇怪技术

Puppeteer 入门

Puppeteer的GitHub链接 本文是对该链接的翻译，扩充解释和举例说明 Puppeteer 是谷歌公司最近推出的基于Node开发的一套高级API库，通过开发协议来控制无界面的浏览器。通俗的说就是有这么一套API, 可以...



乌龟怕铁锤

各种框架文章（先保存下）

完整App@ HackerNews-React-Native用 React Native 完成的 HackerNews 客户端。 WeChat实现类似微信朋友圈或者QQ空间，评论回复，九宫格布局。处理键盘弹出后定位到当前点击的被评论人处。另：滑动时...



wo不懂

iOS开源资源汇总(完整项目,三方,博客,视频)长期更新

下边都学会就大神了: 声明:都是网上搜集的,能标明出处的都标了.别只搜集而不看,与君共勉.. 先看完完整项目完整App@HackerNews-React-Native用 React Native 完成的 HackerNews 客户端。 WeChat实现类似微信朋...



_超

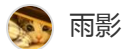
iOS第三方库、插件、大牛博客[OC版]

这是一个用于iOS开发的各种开源库、开源资料、开源技术等等的索引库.转载



【IOS】iOS开发常用三方库、插件、知名博客等等

此文章转自github:<https://github.com/Tim9Liu9/TimLiu-iOS> 介绍 这是一个用于iOS开发的各种开源库、开源资料、开源技术等等的索引库. 具体内容 ===== 版本管理@ 依赖管理@ G...



雨影

我小时候的种种疑点——关于阅读天赋

www.jianshu.com/p/3549871f0007我认为自己也许具有一些阿斯伯格特质。虽然许多舆论对于自疑为阿斯伯格症的人进行无情地鞭答和鄙视，认为这...



鲁叨叨



豆腐鲫鱼汤

一. 谁也想不到我会遇到涂甫。涂甫是一个典型的北方男孩，身型魁梧，五大三粗，毕了业找了份工作，晚上摆烧烤摊儿兼职，操着一口正宗的普通话，“撸烤串儿不大妹子？”。我遇见涂甫的时候是在晚上荒废的...



烧烤摊儿前流连忘返的小王爷

入职随感

2017.10.25 昨天是入职第一天只记得下午跟小伙伴们聊天的时候，我说自己公司访问不了外网，不能上qq、淘宝、京东等一系列网站大家开玩笑说让我辞职吧刹那间，我真的很想立即离开这家公司一个自诩互联网...



susulalala

不是黑暗。而是人生就没有公平。瑞士好。但是中国更好

周末高兴和朋友见面,好久未见我们两个都是酒鬼。喝到12点,服务员说要结束了。我们说好的,马上这点吃的和可爱的酒。万万没想到。她拿来锡纸把我们的吃的全倒在一起,酒倒在塑料杯里。请走。好的我们理解...



2017忘记

我的坏脾气

今天晚上我在和三个朋友一起玩儿游戏。其中一个是我姐姐还有一个叫刘欣怡最后一个叫小坏但是我们闹别扭。就是因为一个游戏，我不跟他们玩儿了，他们就跟我吵架了。我不知道为什么而且他们我说了一些不...



15号小石头聂瑞辰



