1. 请求库的安装
2. Requests： pip install request
3. Selenium：pip install selenium
4. ChromeDriver（谷歌浏览器）：<https://chromedriver.storage.googleapis.com/index.html>

下载完成后，将ChromeDriver的可执行文件放到Python的Scripts目录下。

安装后运行以下代码测试是否安装成功：

from selenium import webdriver

browser=webdriver.Chrome()

1. GeckoDriver（火狐浏览器）：<https://github.com/mozilla/geckodriver/releases>

下载完成后，将GeckoDriver的可执行文件放到Python的Scripts目录下。

安装后运行以下代码测试是否安装成功：

from selenium import webdriver

browser=webdriver.Firefox()

1. PhantomJS：<http://phantomjs.org/download.html>

安装后运行以下代码测试是否安装成功：

from selenium import webdriver

browser=webdriver.PhantomJS()

browser.get("https://www.baidu.com")

print(browser.current\_url)

新版本的Selenium可能会提示报错，但我们可以使用firefox或chrome的headlesss模式,无需重装selenium只需要添加以下代码：

from selenium import webdriver

from selenium.webdriver.chrome.options import Options

chrome\_options = Options()

chrome\_options.add\_argument('--headless')

chrome\_options.add\_argument('--disable-gpu')#上面三行代码就是为了将Chrome不弹出界面，实现无界面爬取

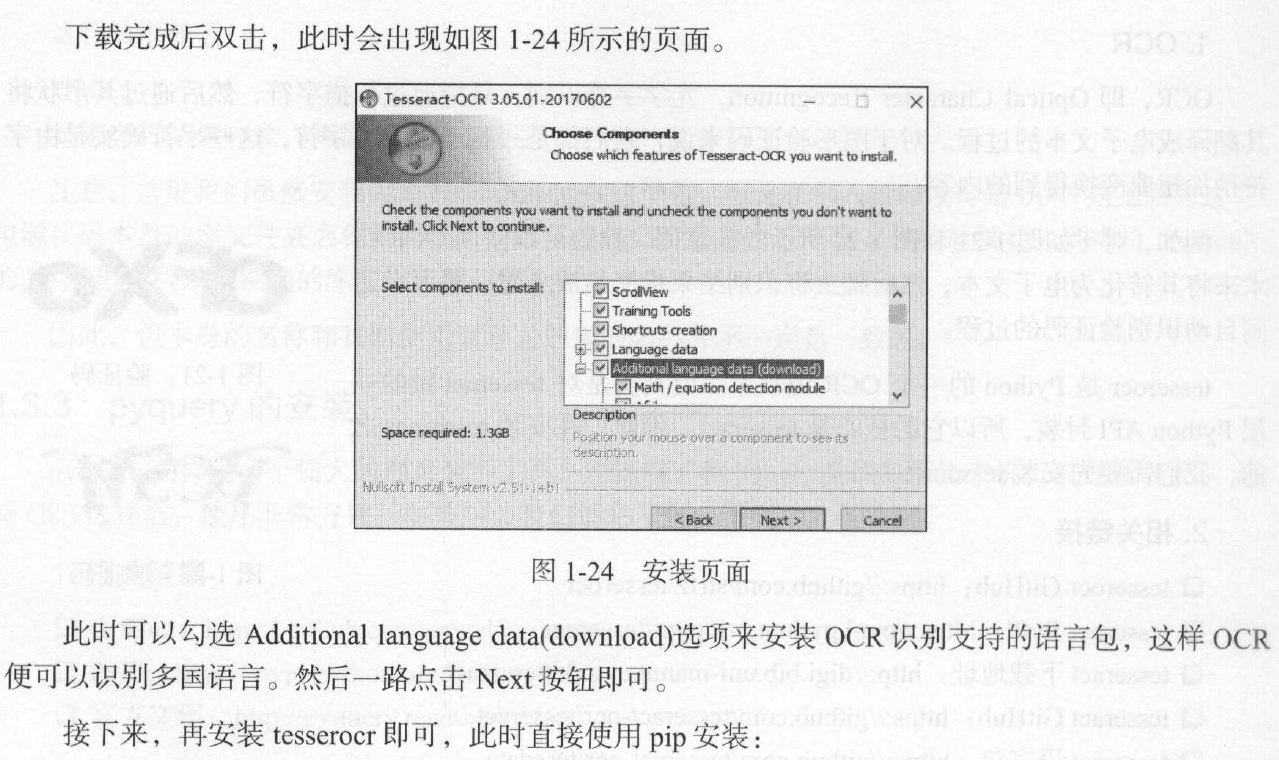
browser = webdriver.Chrome(chrome\_options=chrome\_options)

1. aiohttp：pip install aiohttp （另可选择字符编码检测库cchardet及DNS解析库aiodns）

pip install cchardet aiodns

1. 解析库
2. lxml：pip install lxml
3. beautiful soup：pip install beautifulsoup4
4. pyquery：pip install pyquery
5. tesseracr：<https://digi.bib.uni-mannheim.de/tesseract/>

tesserocr：<https://pypi.org/project/tesserocr/> （pip install tesserocr）



Pip install tesserocr pillow