

Automate your Web Browser With Python (Selenium)

A photograph showing a person's hands holding two mobile phones. The phone on the left is silver and the one on the right is black. In the background, a white computer monitor and a white keyboard are visible on a wooden desk. The text 'FRANK ANEMAET' is overlaid in large, white, outlined letters at the bottom of the image.

FRANK
ANEMAET

Automate your Browser with Python

Frank Anemaet

This book is for sale at <http://leanpub.com/automateyourbrowserwithpython>

This version was published on 2020-08-16



This is a [Leanpub](#) book. Leanpub empowers authors and publishers with the Lean Publishing process. [Lean Publishing](#) is the act of publishing an in-progress ebook using lightweight tools and many iterations to get reader feedback, pivot until you have the right book and build traction once you do.

© 2020 Frank Anemaet

Contents

Introduction	1
Chapter 1	2
Setup	3
Install on Ubuntu Linux	4
Install on Mac OS X	7
Install on Windows	9
Chapter 2	10
Basics	10
Selenium browser	10
Navigate	10
Take screenshot automatically	10
Click button	11
HTML Text input	11
Get HTML element text	12
Scroll page	12
Test if webpage contains text	12
Chapter 3	13
Locate elements by XPath	13
Get all links from webpage	13
Chapter 4	14
Get Links from Webpage	14
Radiobuttons	14
Dropdown	14
Checkbox	14
Table	14
css selector	15
Wait for page load	15
Chapter 5	16
Maximize window	16
Minimize window	16

CONTENTS

Hide window	16
Window size and position	16
Tabs	16
Private mode	16

Introduction

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

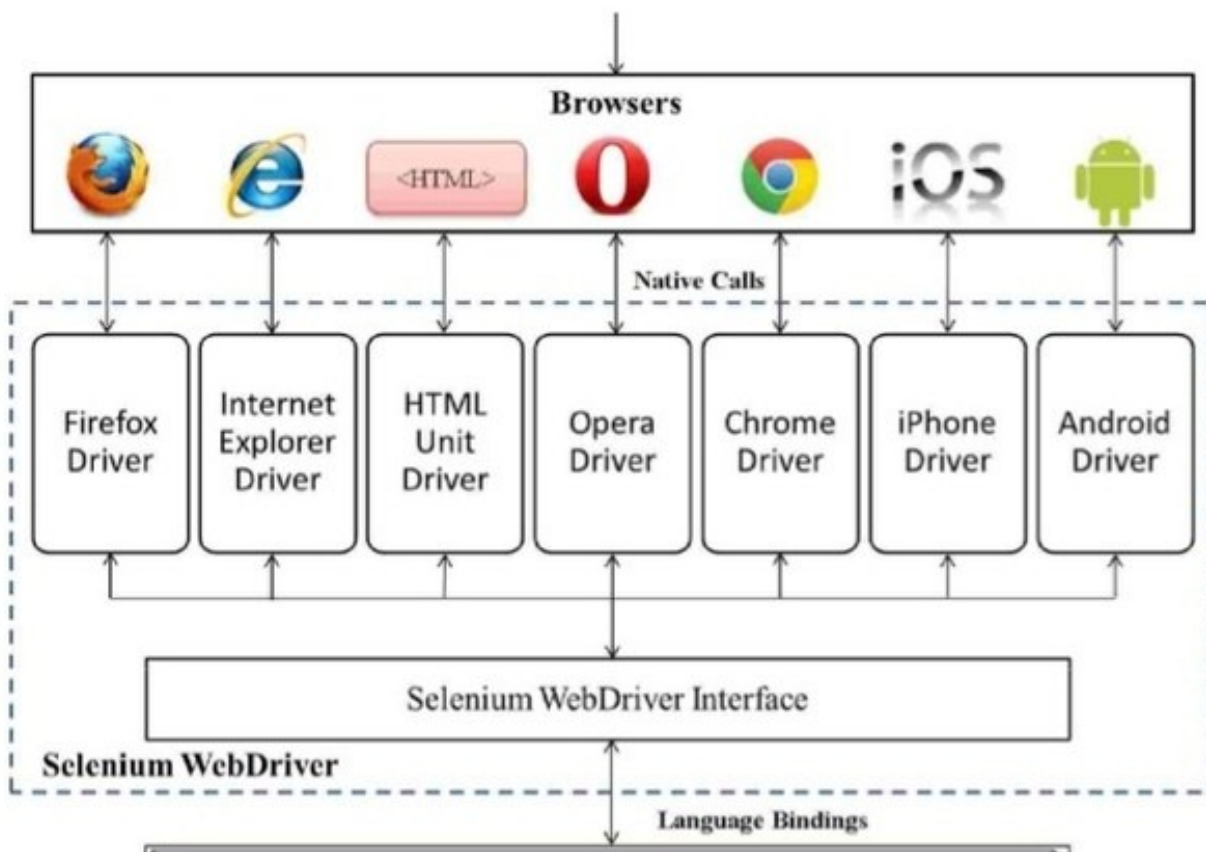
Chapter 1

First lets talk about the basic setup for Python web automation. So you start with Python code. The codes will use a web driver and depending on which browser you want to automate, you need a different driver.

So for Chrome there's the *chromedriver*, for Firefox there's the *geckodriver* and some on.

A webdriver controls a web browser. So depending on the driver it controls Firefox, Chrome, Opera etcetera.

So you are not controlling the web browser directly but you control the Web browser via a *Web Driver*



Control browser via WebDriver

Setup

You need to install the *selenium module* and a *web driver*.

I will show you how to do that. The first thing to do is to *install the selenium module for python*.

You can do that with the command **pip install selenium** or **pip3 install selenium** depending on your operating system.

Most of the times it's `pip install selenium`.

So by typing the commands it will install the selenium module.

```
1 pip install selenium
```

You may want to do this in a *virtual environment (virtualenv)*.

But this is not the only step you need to do.

You also need to install a web driver to install a **web driver**.

The Google Chrome and Chromium, the chrome driver is available on <https://chromedriver.chromium.org/>

For Mozilla Firefox, you can fetch it from <https://github.com/mozilla/geckodriver/releases/tag/v0.26.0> (scroll to bottom).

In the next paragraphs, we'll discuss an easy way to install selenium for the most popular operating systems.

Install on Ubuntu Linux

First make sure you have the pip package manager installed, on Ubuntu this maybe called pip3. If it's not installed you can use apt-get to install it

```
1 sudo apt-get -y install python3-pip
```

For Ubuntu Linux, you can follow the steps below.

```
1 pip3 install selenium
```

You must install the selenium driver too.

Firefox

Download and install the Firefox browser. The next thing you want to do is to install the driver. You can do that manually by downloading the Firefox driver from <https://github.com/mozilla/geckodriver/releases/tag/v> (change to your version) . Then extract it, copy it to /usr/local/bin and make it executable using the command `chmod +x`.

Alternatively, copy the contents of the script below and save it as `firefox.sh`

```
1 echo "Linux script, for Mac OS install geckodriver manually."
2 echo ""
3 echo "Checking latest version..."
4 wget -q https://github.com/mozilla/geckodriver/releases/latest
5 version=$(cat latest | sed -n -e 's!.*<title>Release \(.*\)</title>.*!\1!p')
6 version=$(echo $version | sed 's/ .*//')
7 echo $version
8
9 echo "Installing geckodriver..."
10 wget https://github.com/mozilla/geckodriver/releases/download/v$version/geckodriver-
11 v$version-linux64.tar.gz
12 tar -xvzf geckodriver-v$version-linux64.tar.gz
13 rm geckodriver-v$version-linux64.tar.gz
14 chmod +x geckodriver
15 cp geckodriver /usr/local/bin/
```

Make it executable with the command below

```
1  chmod 755 firefox.sh
2  ./firefox.sh
```

Then type this command to install the Firefox driver. It will install the web driver for you.

Chrome

If you want to use Chrome, install the Chrome web browser and the driver. You can download it from <https://chromedriver.storage.googleapis.com/>

Then extract it and install it.

You can use the chrome install script below, save it as `chrome.sh`:

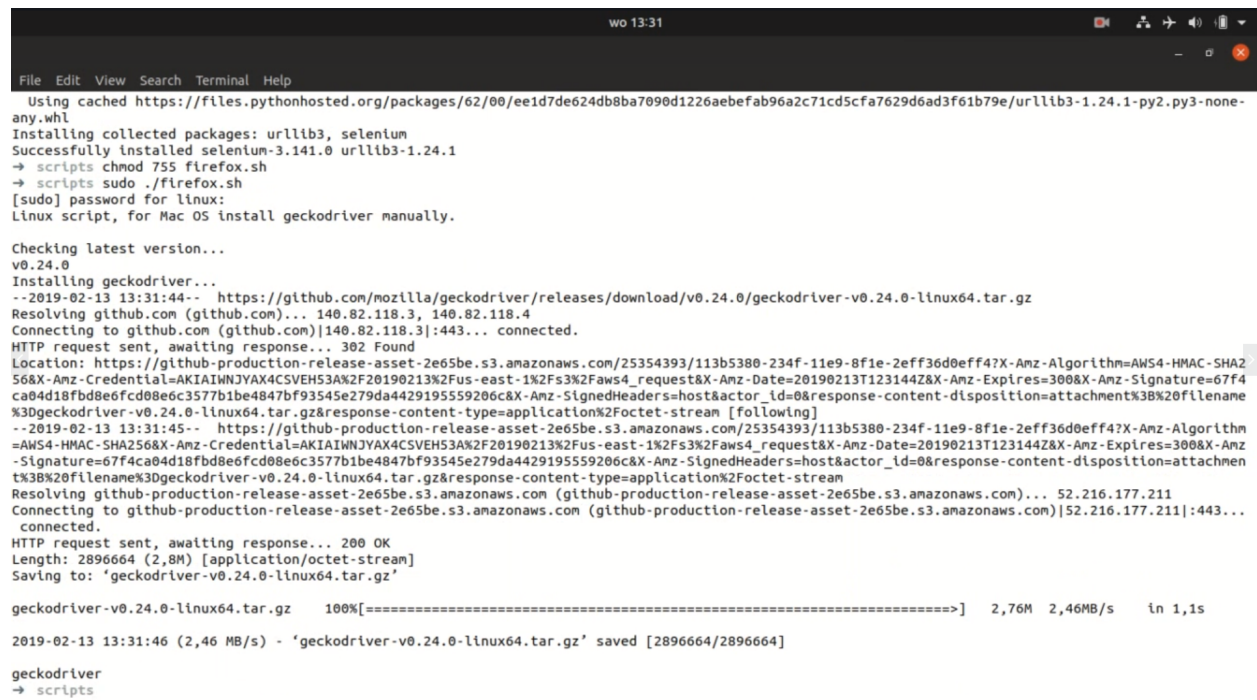
```
1 LATEST=$(wget -q -O - https://chromedriver.storage.googleapis.com/LATEST_RELEASE)
2 wget https://chromedriver.storage.googleapis.com/$LATEST/chromedriver_linux64.zip
3 unzip chromedriver_linux64.zip && sudo ln -s $PWD/chromedriver /usr/local/bin/chrome\
4 driver
```

Then run these commands:

```
1 chmod 755 chrome.sh
2 ./chrome.sh
```

If these commands ran successfully, the **selenium module** and the **webdriver** are installed.

There are the scripts *test-firefox.py* and *test-chrome* to test if installation was successful.



```
wo 13:31
File Edit View Search Terminal Help
Using cached https://files.pythonhosted.org/packages/62/00/ee1d7de624db8ba7090d1226aebefab96a2c71cd5cfa7629d6ad3f61b79e/urllib3-1.24.1-py2.py3-none-any.whl
Installing collected packages: urllib3, selenium
Successfully installed selenium-3.141.0 urllib3-1.24.1
→ scripts chmod 755 firefox.sh
→ scripts sudo ./firefox.sh
[sudo] password for linux:
Linux script, for Mac OS install geckodriver manually.

Checking latest version...
v0.24.0
Installing geckodriver...
--2019-02-13 13:31:44-- https://github.com/mozilla/geckodriver/releases/download/v0.24.0/geckodriver-v0.24.0-linux64.tar.gz
Resolving github.com (github.com)... 140.82.118.3, 140.82.118.4
Connecting to github.com (github.com)[140.82.118.3]:443... connected.
HTTP request sent, awaiting response... 302 Found
Location: https://github-production-release-asset-2e65be.s3.amazonaws.com/25354393/113b5380-234f-11e9-8f1e-2eff36d0eff4?X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIAIWNJYAX4CSVEH53A%2F20190213%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Date=20190213T123144Z&X-Amz-Expires=300&X-Amz-Signature=67f4ca04d18fbd8e6fcd08e6c3577b1be4847bf93545e279da4429195559206c&X-Amz-SignedHeaders=host&actor_id=0&response-content-disposition=attachment%3B%20filename%3Dgeckodriver-v0.24.0-linux64.tar.gz&response-content-type=application%2Foctet-stream [following]
--2019-02-13 13:31:45-- https://github-production-release-asset-2e65be.s3.amazonaws.com/25354393/113b5380-234f-11e9-8f1e-2eff36d0eff4?X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIAIWNJYAX4CSVEH53A%2F20190213%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Date=20190213T123144Z&X-Amz-Expires=300&X-Amz-Signature=67f4ca04d18fbd8e6fcd08e6c3577b1be4847bf93545e279da4429195559206c&X-Amz-SignedHeaders=host&actor_id=0&response-content-disposition=attachment%3B%20filename%3Dgeckodriver-v0.24.0-linux64.tar.gz&response-content-type=application%2Foctet-stream
Resolving github-production-release-asset-2e65be.s3.amazonaws.com (github-production-release-asset-2e65be.s3.amazonaws.com)... 52.216.177.211
Connecting to github-production-release-asset-2e65be.s3.amazonaws.com (github-production-release-asset-2e65be.s3.amazonaws.com)[52.216.177.211]:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 2896664 (2,8M) [application/octet-stream]
Saving to: 'geckodriver-v0.24.0-linux64.tar.gz'

geckodriver-v0.24.0-linux64.tar.gz  100%[=====]  2,76M  2,46MB/s   in 1,1s

2019-02-13 13:31:46 (2,46 MB/s) - 'geckodriver-v0.24.0-linux64.tar.gz' saved [2896664/2896664]

geckodriver
→ scripts
```

Installation output on Ubuntu Linux

Install on Mac OS X

If you use Apple Mac OS X, you can install selenium and the web driver too. The process is similar to installation on Linux.

First open a terminal.

Firefox

Create this script and save it as firefox.sh

```
1  echo "Web Driver install for Mac OS X (Firefox, Geckodriver)."
2  echo ""
3  echo "Checking latest version..."
4  wget -q https://github.com/mozilla/geckodriver/releases/latest
5  version=$(cat latest | sed -n -e 's!.*<title>Release \(.*\)</title>.*!\1!p')
6  version=$(echo $version | sed 's/ .*//')
7  echo $version
8
9  echo "Installing geckodriver..."
10 wget https://github.com/mozilla/geckodriver/releases/download/$version/geckodriver-$\
11 version-macos.tar.gz
12 tar -xvzf geckodriver-$version-macos.tar.gz
13 rm geckodriver-$version-macos.tar.gz
14 chmod +x geckodriver
15 cp geckodriver /usr/local/bin/
```

To install Selenium Firefox for Mac OS X:

```
1  pip3 install selenium
2  chmod 755 firefox.sh
3  sudo ./firefox.sh
4  python3 test-firefox.py
```

Chrome

For Chrome it's similar. You should install the driver and the selenium module. If you don't have Chrome installed, you should install the browser too.

First install the selenium module

```
1 pip3 install selenium
```

Next the driver needs to be installed.

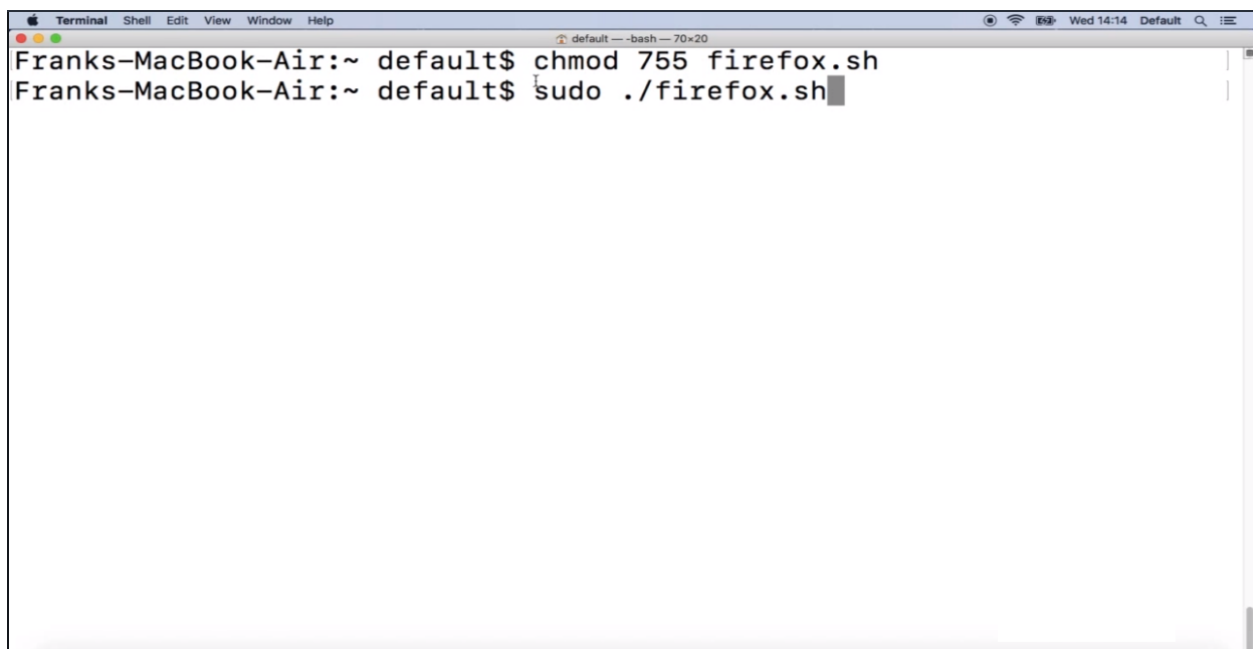
Copy the contents below and save it as chrome.sh

```
1 LATEST=$(wget -q -O - https://chromedriver.storage.googleapis.com/LATEST_RELEASE)
2 wget https://chromedriver.storage.googleapis.com/$LATEST/chromedriver_mac64.zip
3 unzip chromedriver_mac64.zip && sudo ln -s $PWD/chromedriver /usr/local/bin/chromedr\
4 iver
```

To install the Chrome version on Mac OS X use:

```
1 chmod 755 chrome.sh
2 sudo ./chrome.sh
```

This will install both the **web driver** and the **selenium module**.



Installation output on Mac OS X

Install on Windows

To install Selenium on Windows 10:

1. Install Python
2. Install Latest Web Browser (Firefox, Chrome, Chromium...)
3. Download Web Driver from <https://www.seleniumhq.org/download/>
4. Add drivers path to environment variable PATH

When using the program, you can set the path to the driver like this:

```
1 from selenium import webdriver
2
3 options = webdriver.ChromeOptions()
4 driver = webdriver.Chrome(executable_path="path/to/chromedriver.exe", chrome_options\
5 =options)
```

Chapter 2

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Basics

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Selenium browser

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Navigate

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Mimick browsing behavior

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Back, refresh

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Take screenshot automatically

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Get HTML Source

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Driver variables

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Take bulk screenshots

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Full page screenshot

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Click button

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

HTML Text input

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Textfield input

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Textbox input

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Get HTML element text

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Scroll page

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Test if webpage contains text

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Chapter 3

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Locate elements by XPath

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

XPath in Firefox

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Xpath in Chrome

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Get all links from webpage

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Chapter 4

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Get Links from Webpage

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Radiobuttons

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Dropdown

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Checkbox

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Table

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Count rows in table

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Table to list

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

css selector

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Wait for page load

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Check if element exists

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Find by css selector

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Find by id

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Find by xpath

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Chapter 5

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Maximize window

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Minimize window

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Hide window

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Window size and position

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Tabs

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Private mode

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Private mode in Firefox

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.

Private mode in Chrome

This content is not available in the sample book. The book can be purchased on Leanpub at <http://leanpub.com/automateyourbrowserwithpython>.