

ECON446 : Applications of Cloud Computing and Big Data

Week 3, Lecture 3 | Selenium Intro

Sam Borghese

Tuesday, April 15th, 2024

1. What is Selenium?
2. Problems selenium can fix
3. What Selenium can do
4. Examples

What is Selenium?

Selenium

Automate Web Browsers

Free Open source automated testing framework used
to validate web applications across different browsers
and platforms

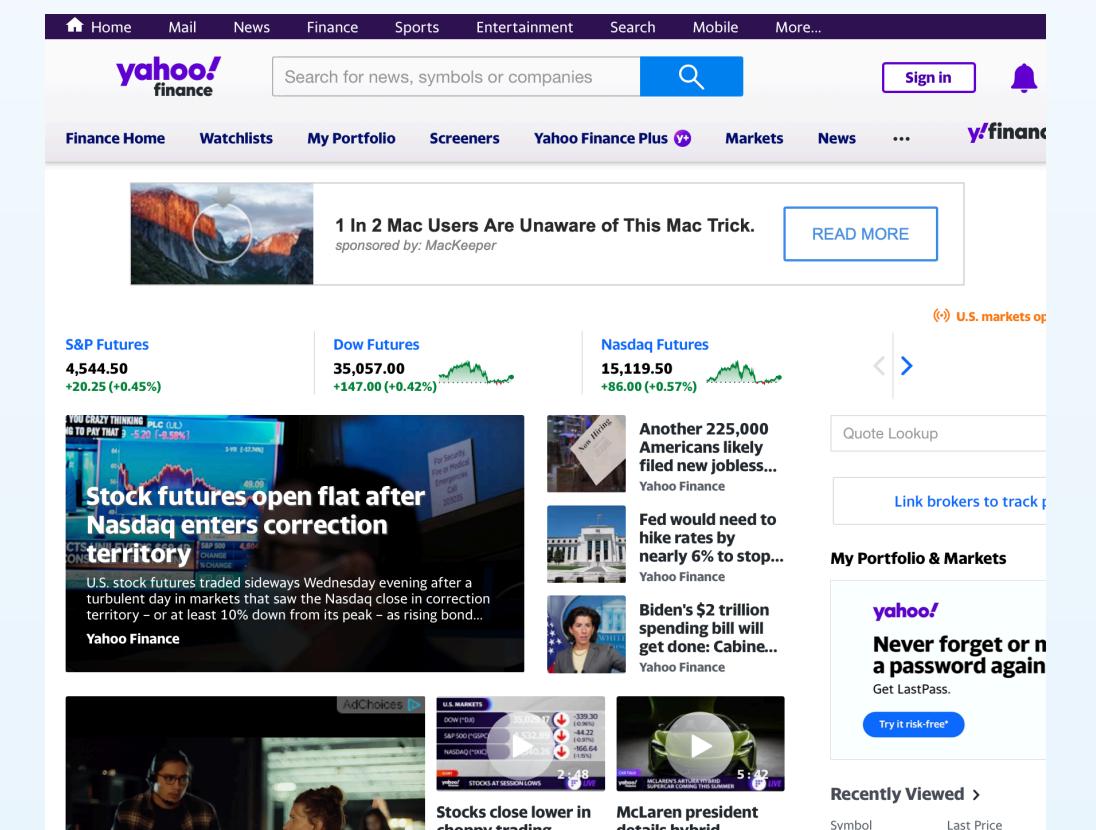
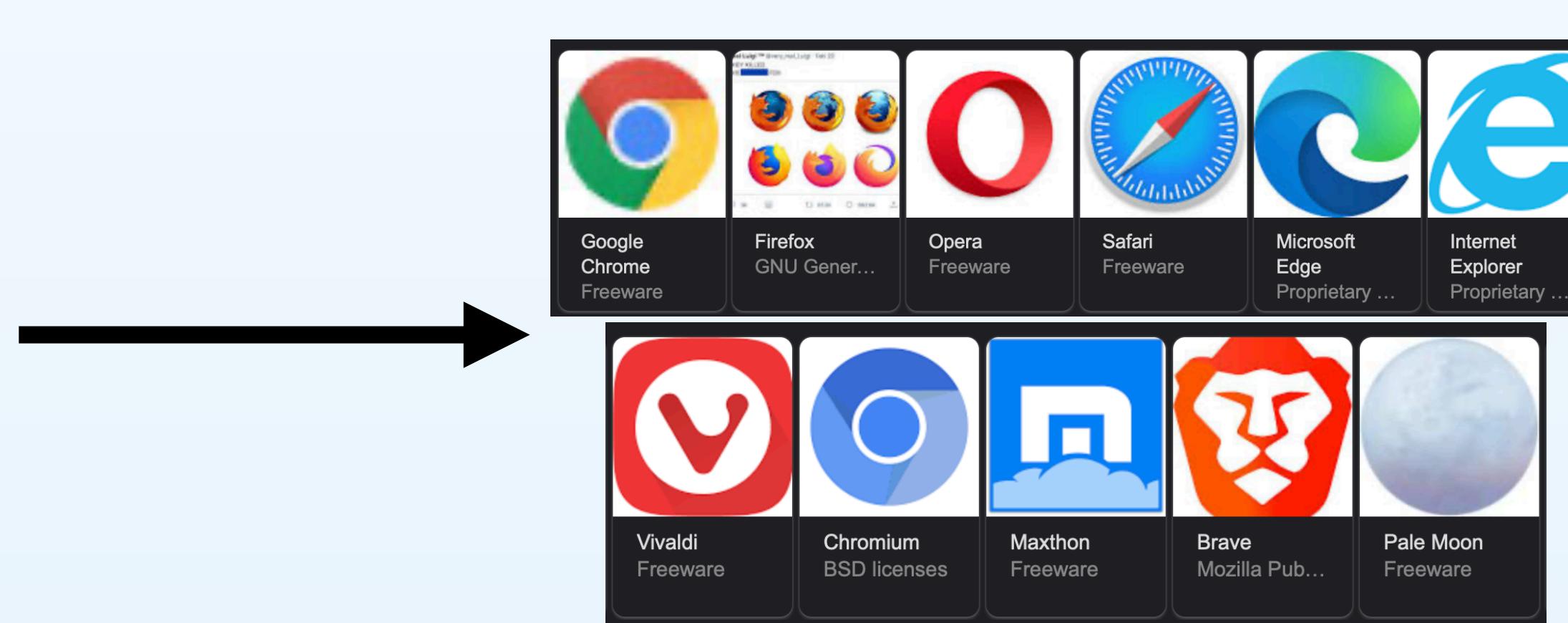


The name Selenium comes from a joke made by Huggins in an email, mocking a competitor named Mercury, saying that you can cure mercury poisoning by taking selenium supplements.

Selenium

Test across many Web browsers at rapidly

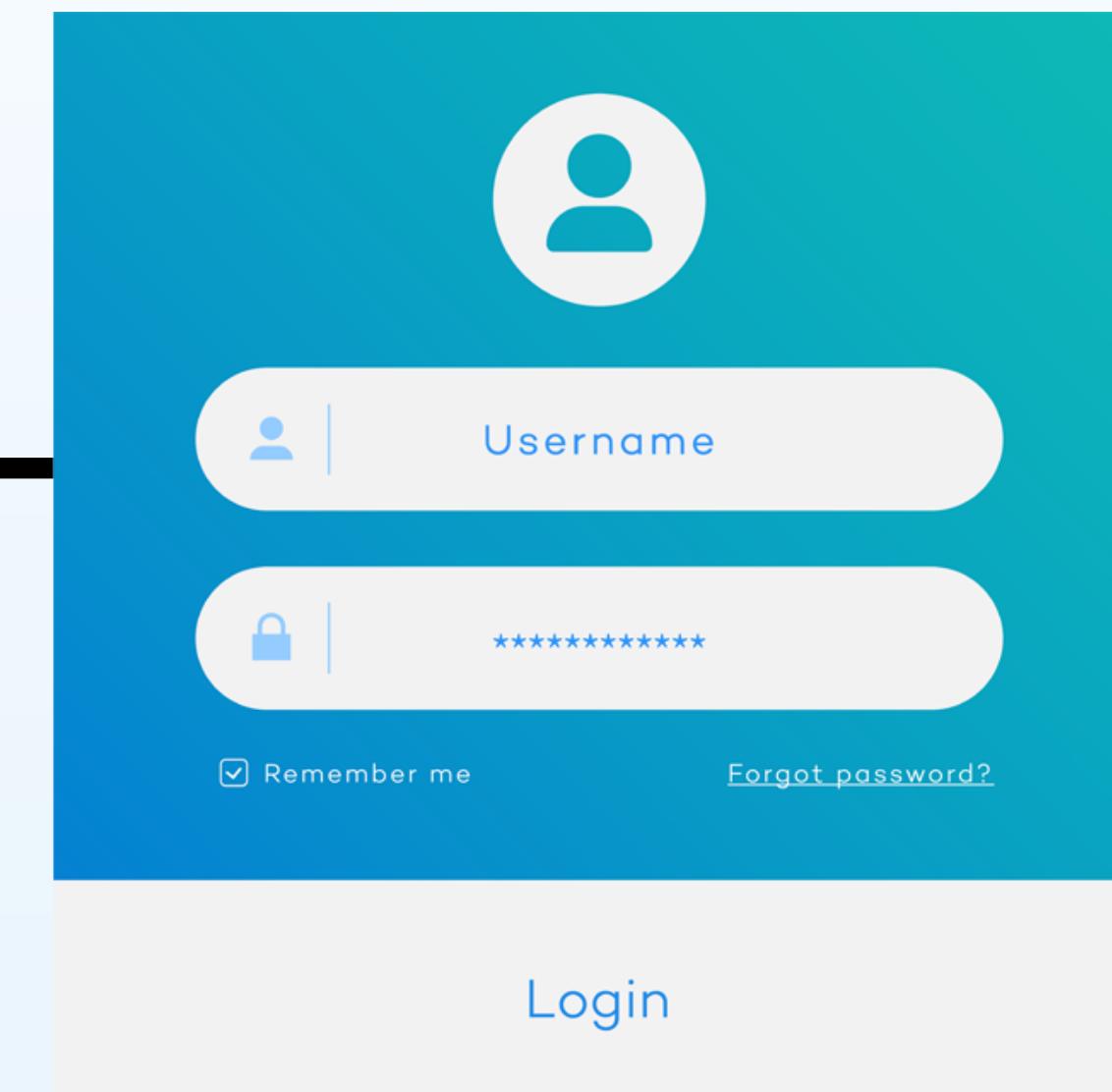
1. Create a WebDriver instance.
2. Navigate to a webpage.
3. Locate a web element on the webpage via locators in selenium.
4. Perform one or more user actions on the element.
5. Preload the expected output/browser response to the action.
6. Run test.



Selenium Web Driver Test Example

Selenium is Used to test for a specific output across different Browsers

```
username.sendKeys("abc@gmail.com");  
password.sendKeys("your_password");
```



```
String expectedUrl= driver.getCurrentUrl();
```

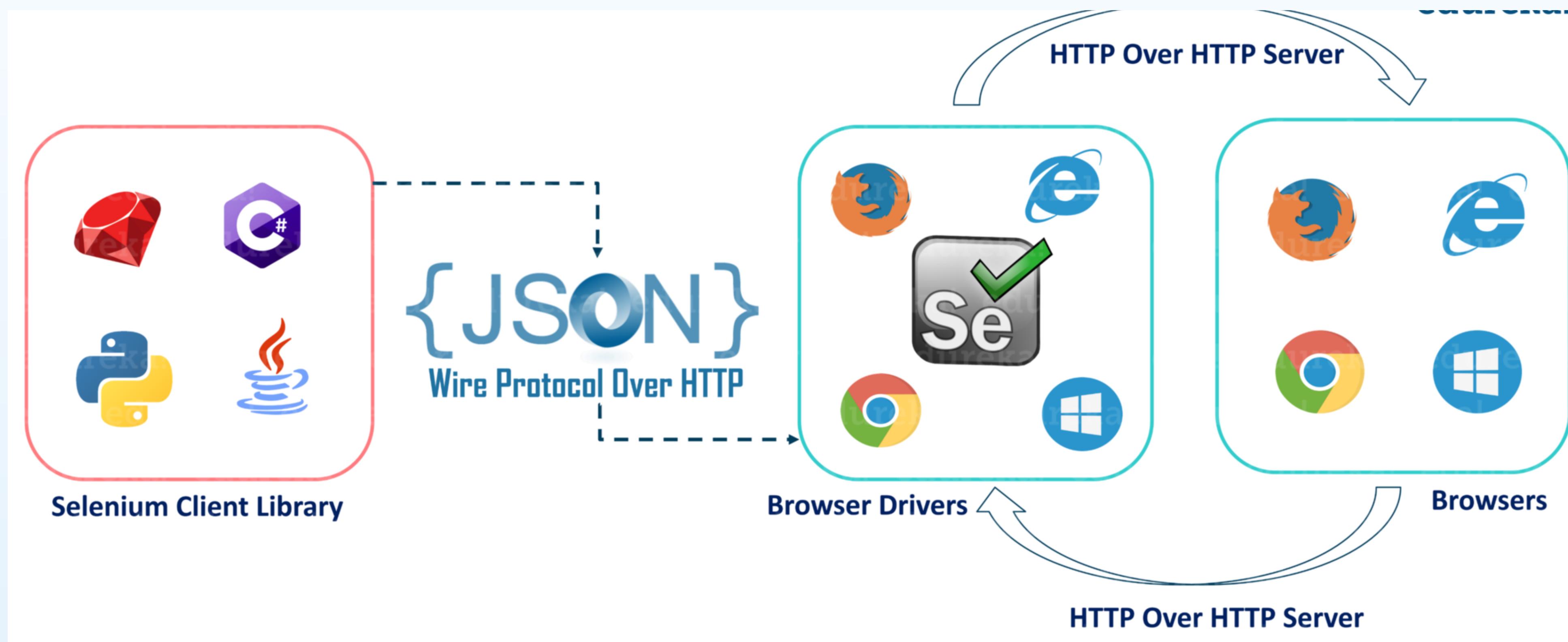
?=

```
String actualUrl="https://live.browserstack.com/dashboard";
```

Web Drivers / Browser Drivers

Selenium uses a host of Browser Drivers to interact with browsers

Web Driver : A browser automation framework that works by using a collection of APIs

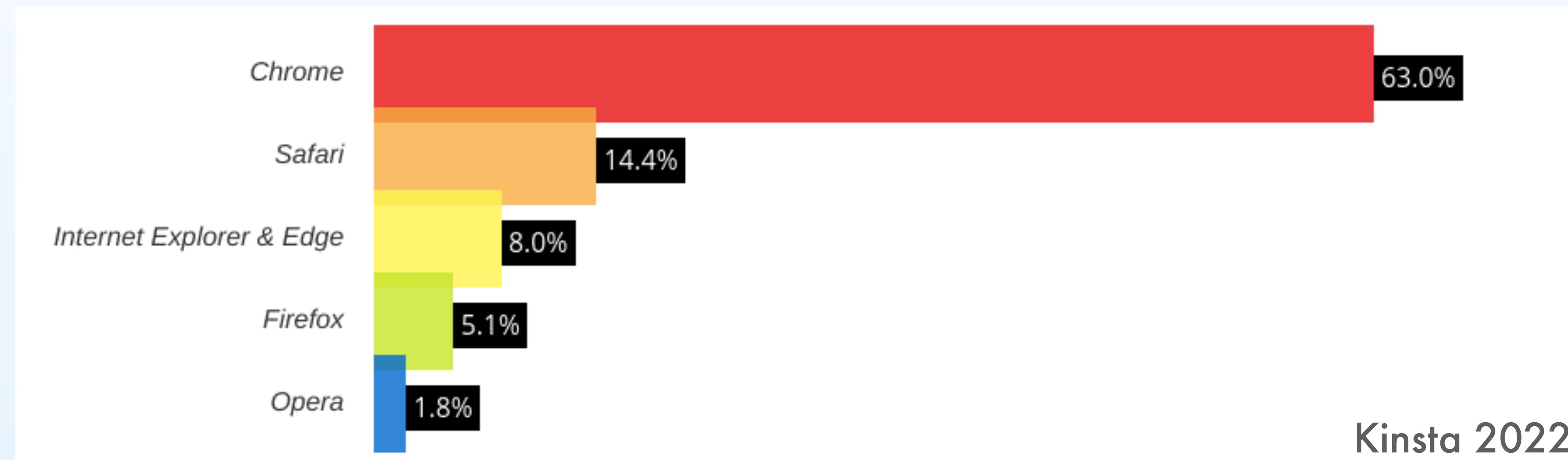


Selenium Driver Website

Download Chrome Driver

<https://sites.google.com/chromium.org/driver/>

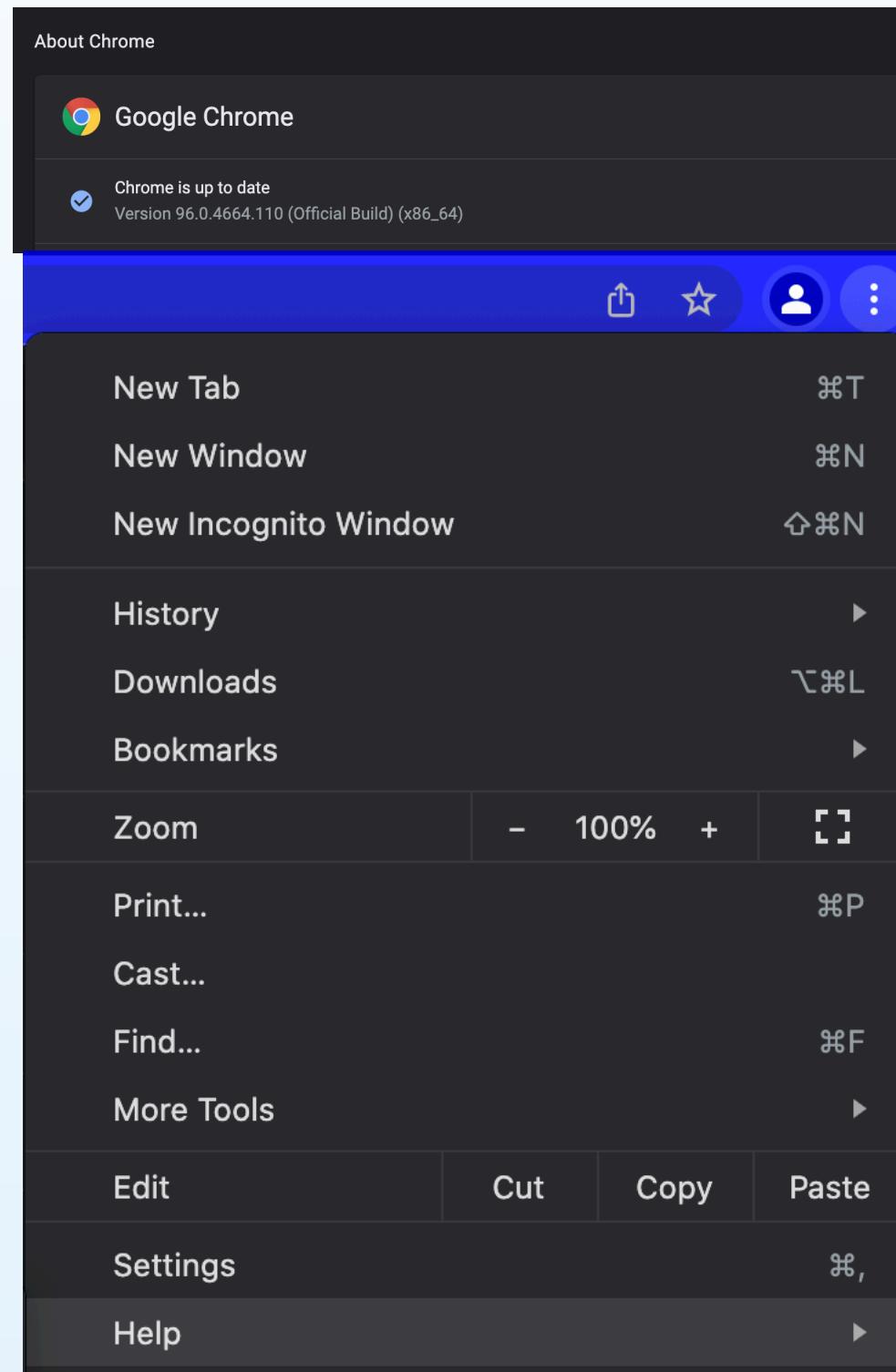
We'll use Chrome Driver because it is the most popular



Selenium Driver Website

<https://sites.google.com/chromium.org/driver/>

Check Chrome Version/Update



Download Chrome Driver

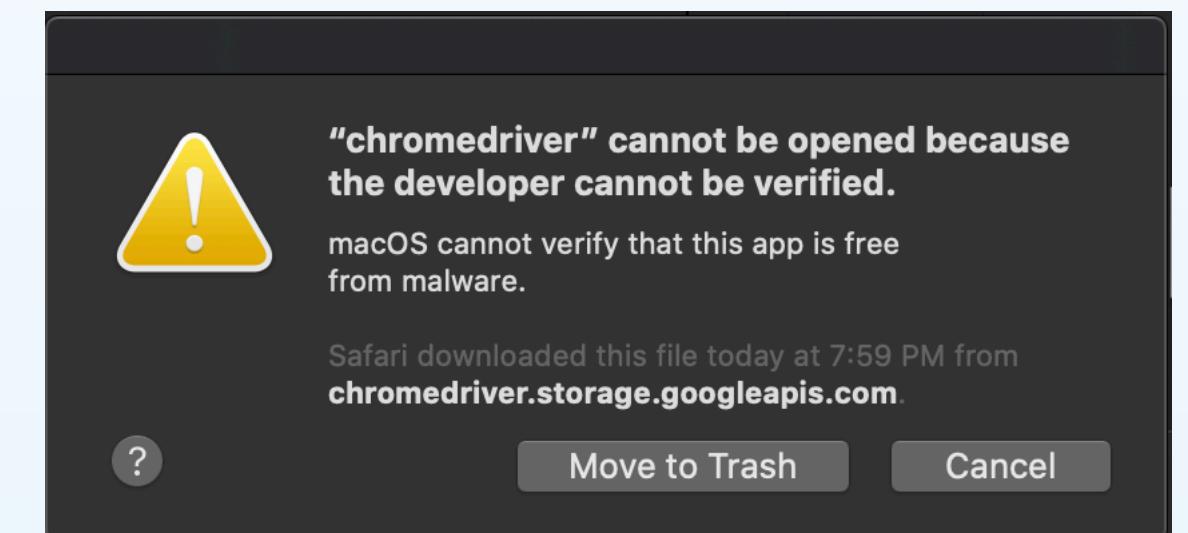
A screenshot of a file download interface, likely from a web browser, showing a list of files:

Name	Last modified	Size	ETag
Parent Directory		-	-
chromedriver_linux64.zip	2021-11-16 09:35:54	9.64MB	318c85b6709902b2e62a55da22362b05
chromedriver_mac64.zip	2021-11-16 09:35:56	7.85MB	60622681d95f2087005385628cea2d1a
chromedriver_mac64_m1.zip	2021-11-16 09:35:58	7.44MB	8dd9b4b86d433e51a0b04089c6c13fc
chromedriver_win32.zip	2021-11-16 09:36:00	5.80MB	b1c93546ea5f304a9e7ac8b3944af7dc
notes.txt	2021-11-16 09:36:05	0.00MB	7e1ce9231396cabd06e1ff2b0e539fe5

All versions available in [Downloads](#)

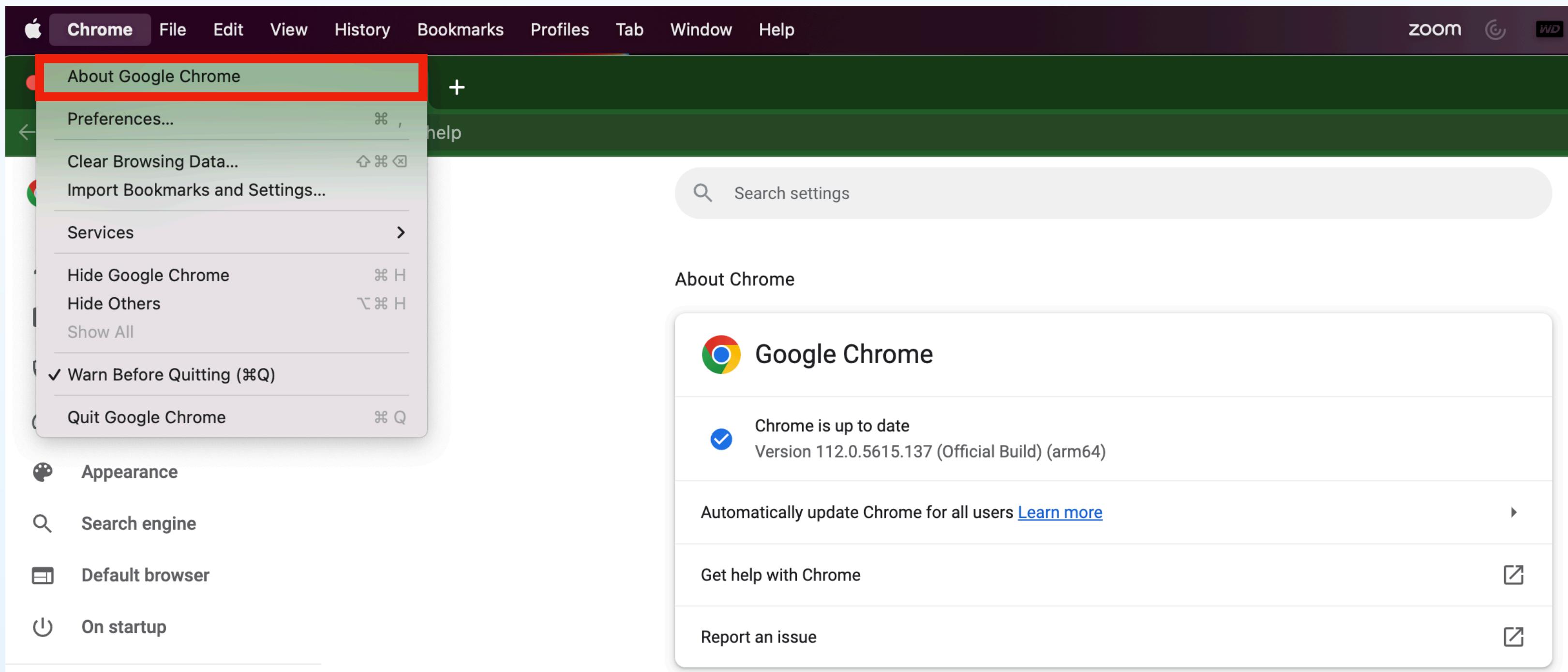
- Latest beta release: [ChromeDriver 97.0.4692.36](#)
- Latest stable release: [ChromeDriver 96.0.4664.45](#)

Allow Chrome Driver



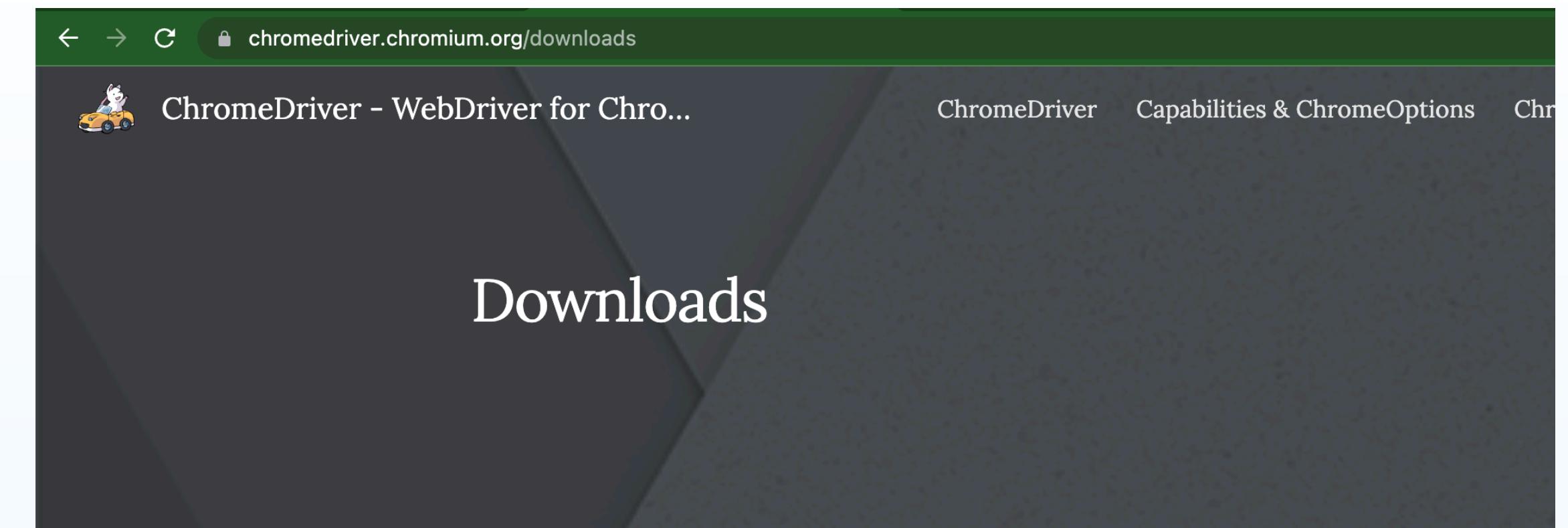
Find Chrome Version

Chrome Version



Find Chrome Version

Download corresponding Driver



Current Releases

- If you are using Chrome version 113, please download [Chromedriver 113.0.5672.24](#)
- If you are using Chrome version 112, please download [Chromedriver 112.0.5615.49](#)
- If you are using Chrome version 111, please download [Chromedriver 111.0.5563.64](#)
- For older version of Chrome, please see below for the version of Chromedriver that supports it.

If you are using Chrome from Dev or Canary channel, please following instructions on the [Chromedriver Canary](#) page.

For more information on selecting the right version of Chromedriver, please see the [Version Selection](#) page.

[Chromedriver 113.0.5672.24](#)

Supports Chrome version 113

- Resolved issue 4205: Same object ids in Classic and BiDi [Pri-1]
- Resolved issue 4302: Don't assume that Mapper is in the first tab in ExecuteGetWindowHandles [Pri-1]
- Resolved issue 4356: Chrome 110 not utilizing pref value "download.default_directory" [Pri-1]

For more details, please see the [release notes](#).

[Chromedriver 112.0.5615.49](#)

Supports Chrome version 112

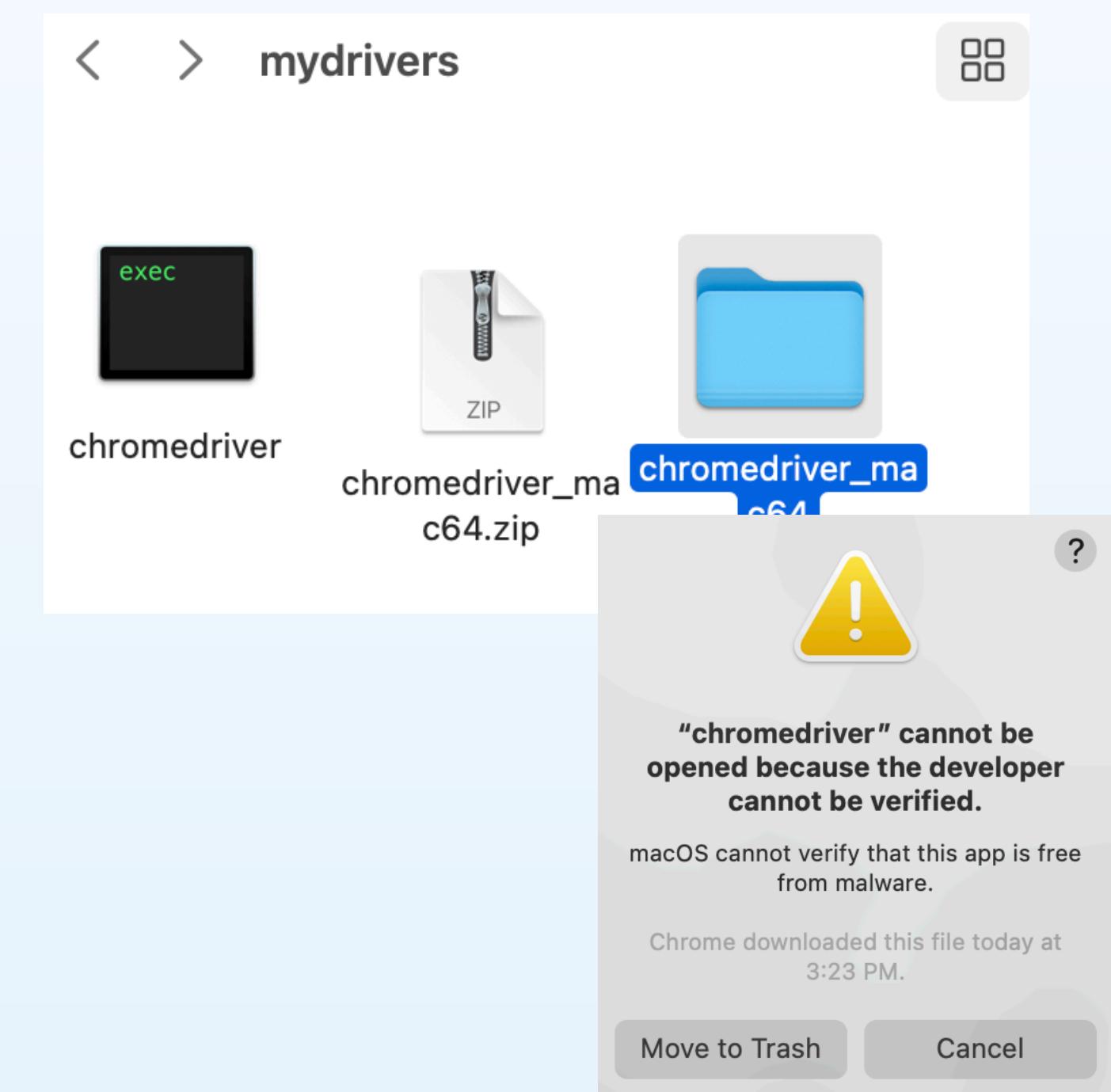
- Resolved issue 3517: Enable print feature for non-headless [Pri-1]
- Resolved issue 4419: Large overhead on Speedometer when using chromedriver [Pri-3]

For more details, please see the [release notes](#).

Store Chromedriver

Download the chrome driver for your machine and store the “chrome driver” somewhere

Index of /112.0.5615.49/				
Name	Last modified	Size	ETag	
Parent Directory		-		
chromedriver_linux64.zip	2023-04-05 10:27:32	6.75MB	9f887638c5e8bc5313d027a802b092cc	
chromedriver_mac64.zip	2023-04-05 10:27:35	8.79MB	a9f0c8afeaecc961d3bdf087f33f71c9	
chromedriver_mac_arm64.zip	2023-04-05 10:27:39	8.04MB	19231ea55d8f84baf1aff6e38c6a22e9	
chromedriver_win32.zip	2023-04-05 10:27:42	6.80MB	3aa0e2c3dca02a93422152248411c964	
notes.txt	2023-04-05 10:27:48	0.00MB	07a5cc857188e777937ca2dcc0017fb6	



Problem Selenium can address

Using Requests Sometimes Gives a Different result than Inspect Element

Inspect Element



The screenshot shows a news article titled "Manchester Beats London as Best Place for New Jobs in England". The article discusses job market trends post-pandemic. The browser's developer tools are open, with the "Elements" tab selected. A specific element, an anchor tag with the class "single-story-module_headline-link", is highlighted with a yellow box. The "Computed" tab in the styles panel shows properties like position, margin, and border.

```
<a href="https://www.bloomberg.com/graphics/2023-reed-jobs-report-manchester-beats-london-new-jobs-in-england/?srnd=premium" class="single-story-module_headline-link">Manchester Beats London as Best Place for New Jobs in England</a> = $0
```

Read through Requests

```
In [15]: 1 soup
Out [15]: <!DOCTYPE html>

<html lang="en" prefix="og: http://ogp.me/ns#">
<head>
<meta charset="utf-8"/>
<script async="" src="https://securepubads.g.doubleclick.net/tag/js/gpt.js"></script>
<title>Bloomberg.com</title><link href="https://www.bloomberg.com/" rel="stylesheet" type="text/css" />
```

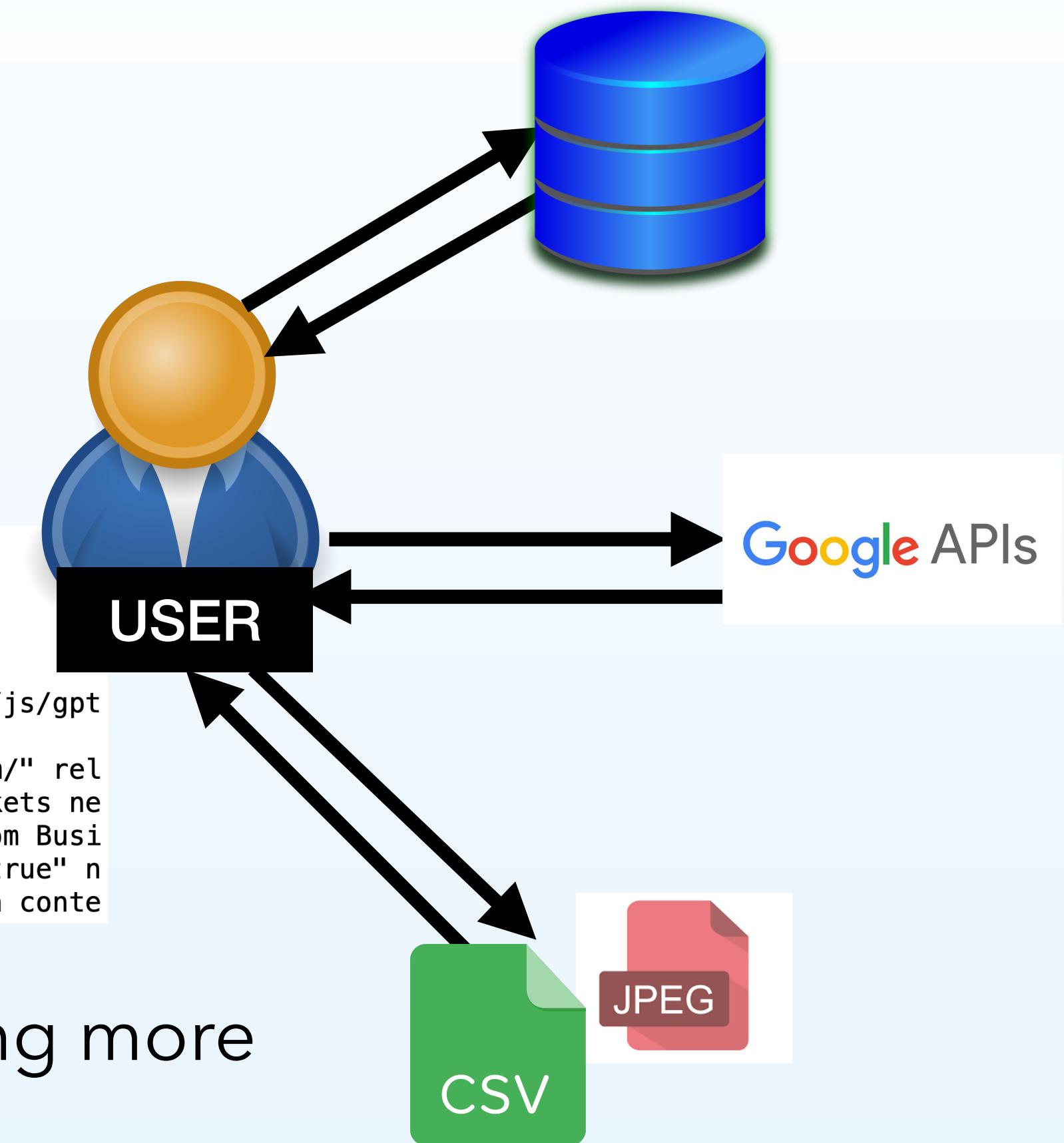
Not found Contains X

Why does this happen?

Javascript

<https://www.bloomberg.com>

```
<!DOCTYPE html>
<html lang="en" prefix="og: http://ogp.me/ns#>
<head>
<meta charset="utf-8"/>
<script async="" src="https://securepubads.g.doubleclick.net/tag/js/gpt.js"></script>
<title>Bloomberg.com</title><link href="https://www.bloomberg.com/" rel="canonical"/><meta content="Bloomberg delivers business and markets news, data, analysis, and video to the world, featuring stories from Businessweek and Bloomberg News" name="description"/><meta content="true" name="homepage"/><meta content="unsafe-url" name="referrer"/><meta conte
```



Javascript can wait for some user interaction before loading more HTML

I.E. : Waiting on the page, scrolling on the page, clicking items, etc.

What Selenium can do

Locating the Web Driver on your computer

Must Locate the Path where the Chrome

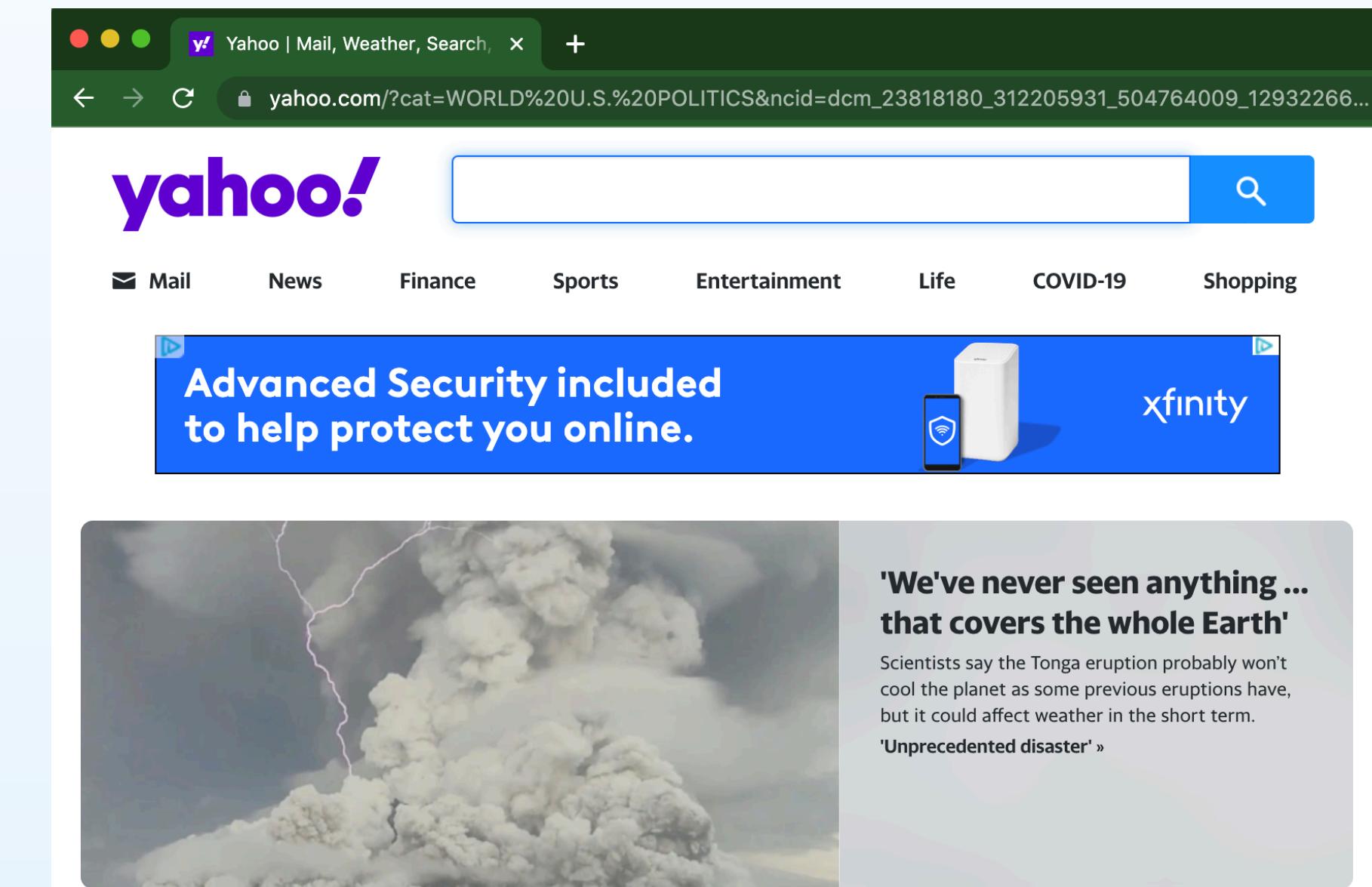
```
import selenium
from selenium import webdriver
from selenium.webdriver.common.keys import Keys
```

```
PATH = "/Users/windows/Desktop/mydrivers/chromedriver"
```

Closing a Browser Window

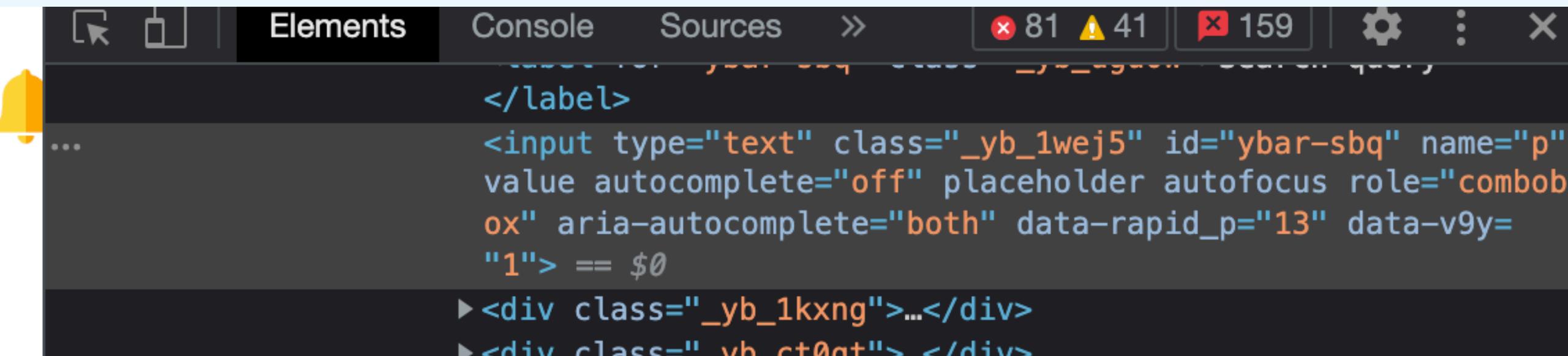
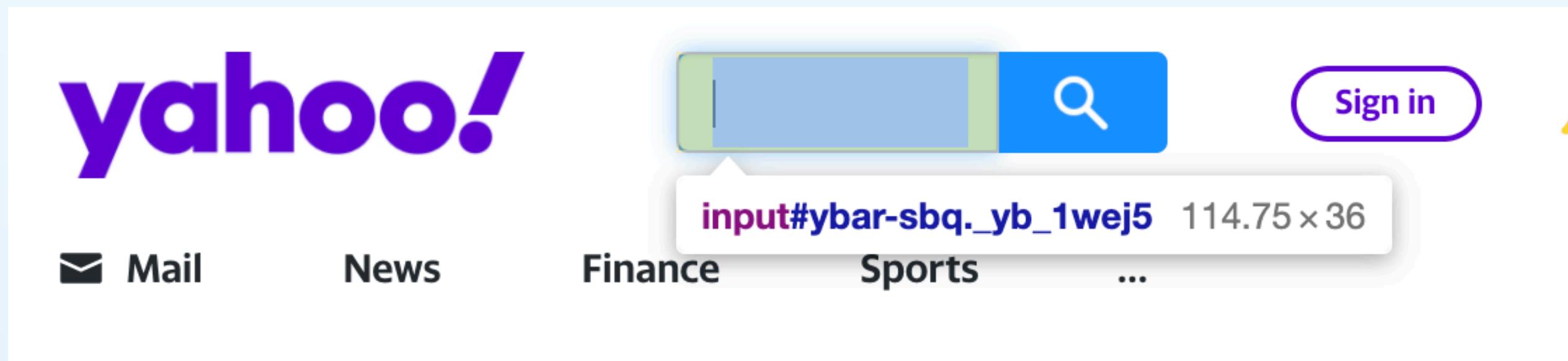
If you only have one tab open then these commands are equivalent

```
# Close The given Tab  
driver.close()  
# Close the entire browser window  
driver.quit()  
# Need to restart driver with driver.get(URL) after quit
```



Finding Elements

```
# p is the name of the search box  
search = driver.find_element_by_name("p")
```

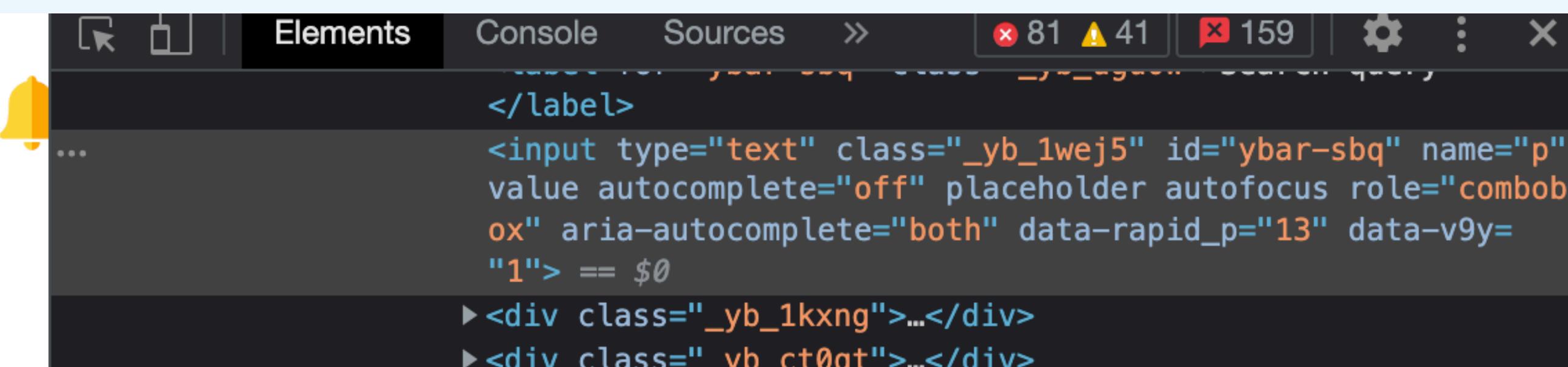
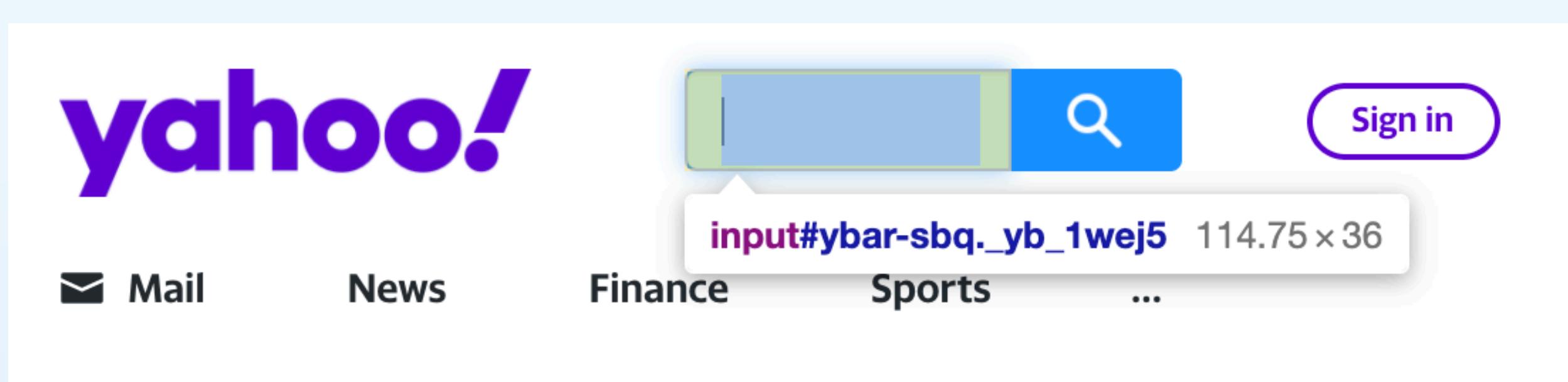


Finding Elements

There are many ways to find elements using Methods

Reminder : Use `find dir()` and `?` To investigate objects

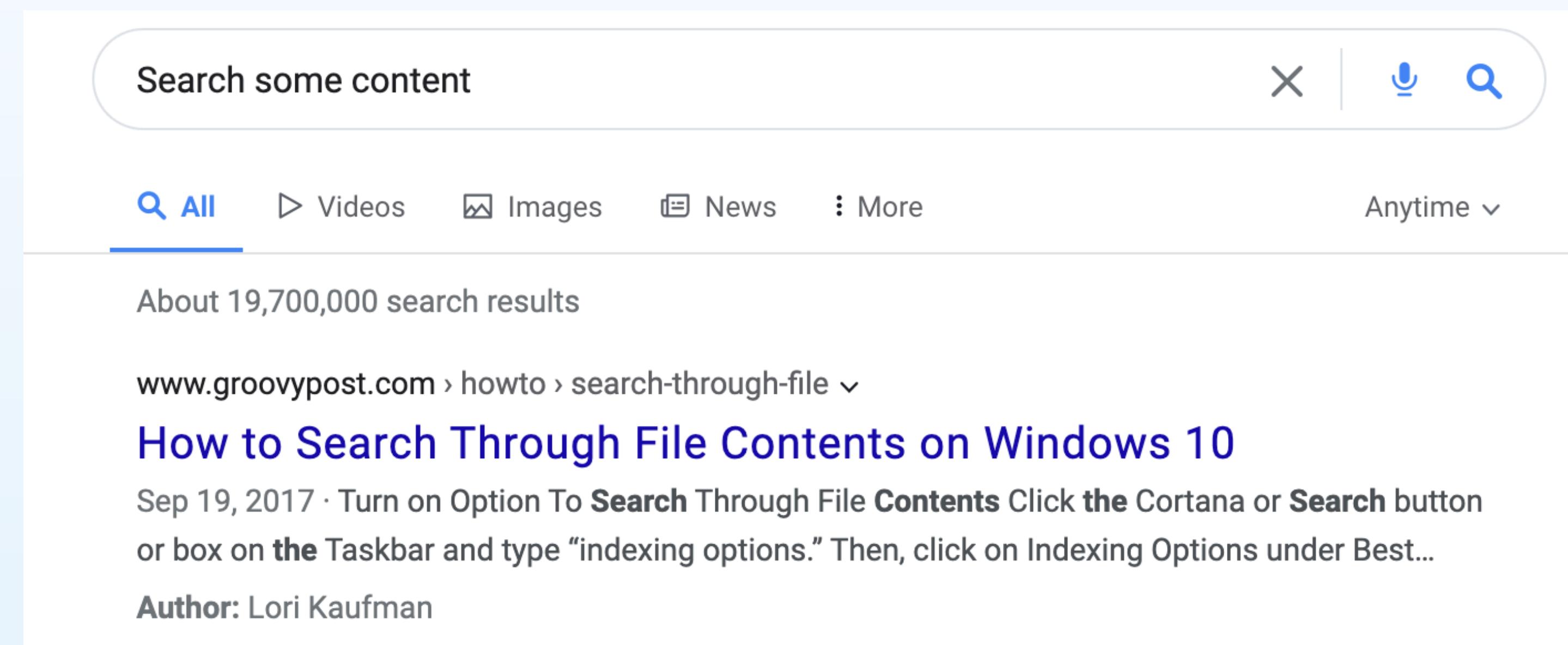
```
'find_element',
'find_element_by_class_name',
'find_element_by_css_selector',
'find_element_by_id',
'find_element_by_link_text',
'find_element_by_name',
'find_element_by_partial_link_text',
'find_element_by_tag_name',
'find_element_by_xpath',
'find_elements',
'find_elements_by_class_name',
'find_elements_by_css_selector',
'find_elements_by_id',
'find_elements_by_link_text',
'find_elements_by_name',
'find_elements_by_partial_link_text',
'find_elements_by_tag_name',
'find_elements_by_xpath',
```



Send Text to Search Boxes

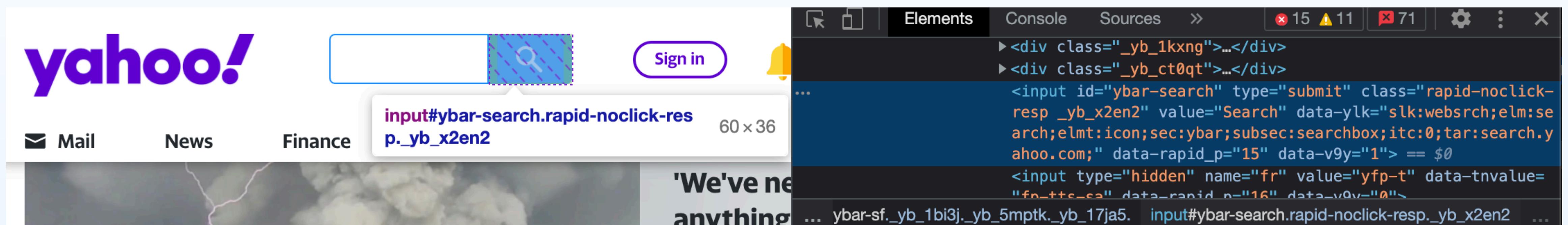
Using `.send_keys()` you can send information to the webpage

```
search.send_keys("Search some content")
search.send_keys(Keys.RETURN)
```



Click Things

Any element is clickable. Some will have a reaction like buttons or hypertext while others won't



```
button = driver.find_element_by_id("vert-search-button")
button.click()
```

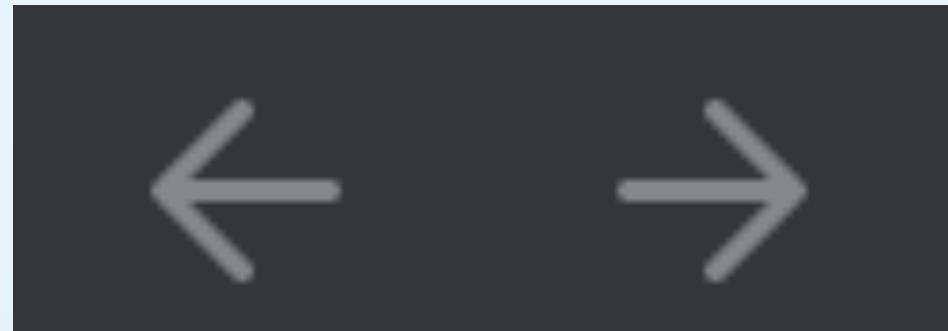
Easier way to Click Hyperlinks

Using “Link Text”

```
link = driver.find_element_by_link_text("Finance")
link.click()
```

Navigate Moving Pages

```
driver.back()
driver.forward()
```



Wait Until Certain Elements Appear

Wait until a certain element has appeared in a page

```
from selenium.webdriver.common.by import By
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
```

Max time to wait (sec)

```
WebDriverWait(driver, 10).until(
    EC.presence_of_element_located((By.LINK_TEXT, "Finance")))
```

Element

Wait for a given amount of time (Not as definite)

```
driver.implicitly_wait(5)
time.sleep(5)
```

Find all Attributes with a Particular tag

Similar to beautiful Soup at this scale

```
h3s = driver.find_elements_by_tag_name('h3')
d  = [h.text for h in h3s]
```

Scroll a Webpage

Scroll to load more of a webpage

```
driver.execute_script("window.scrollTo(0,document.body.scrollHeight)")
```

Add Options

Host of options can be found in the Documentation

<https://selenium-python.readthedocs.io/api.html#>

```
WINDOW_SIZE = "1920,1080"
options = webdriver.ChromeOptions()
options.add_experimental_option("excludeSwitches", ["enable-automation"])
options.add_experimental_option('useAutomationExtension', False)
options.add_argument("--disable-blink-features=AutomationControlled")
options.add_argument("--disable-gpu")
options.add_argument("--no-sandbox")
options.add_argument("--allow-insecure-localhost")
options.add_argument("user-agent=[Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/83.0.4103.106 Safari/537.36]")
options.add_argument("--start-maximized")
options.add_argument("--window-size=%s" % WINDOW_SIZE)
options.add_argument("--headless")
```

Examples

LinkedIn is a Data Company

They protect their data and make our past scraping techniques obsolete

LinkedIn has been very clear about its stance on data scraping—it's against it. The social media giant has even tried to fight data scraping in court. However, the 9th U.S. Circuit Court of Appeals ruled that scraping public data from websites is not prohibited by any federal law.

Scrape information from Disney

Example 1 :

Open a browser using selenium, using the driver search for Disney jobs in the United States.

Find the number of new jobs posted

Example 2 :

Get a list of all the positions on the front page. Hint : You will have to scroll down to load them all