



kentavr009

10 ноя в 10:48

Решение reCAPTCHA в Selenium на полном автомате (адаптация англоязычной статьи)

Простой

4 мин

6.4K

Python*

Кейс

Делаю робкие попытки в автоматизации и столкнулся с часто возникающей проблемой, распознавание Рекапчи (reCapt cha).

Естественно, я понимаю что по данной теме написано много гайдов, мануалов и статей, но согласитесь - интересно описать собственный опыт.

Итак, за основу я взял англоязычный мануал, который попался мне на глаза буквально пару дней назад и я решил его потестить (а так как написан он сервисом по распознаванию капчи, который я и использую, то почему нет - кстати, ребята из 2капча - принимаю благодарность в виде зеленых бумажек, если интересно))))

Итак, не будем долго и нудно расписывать, зачем и почему нужен Селениум, как это важно для автоматизации и кога он возник и кто его придумал, переходим к самой сути:

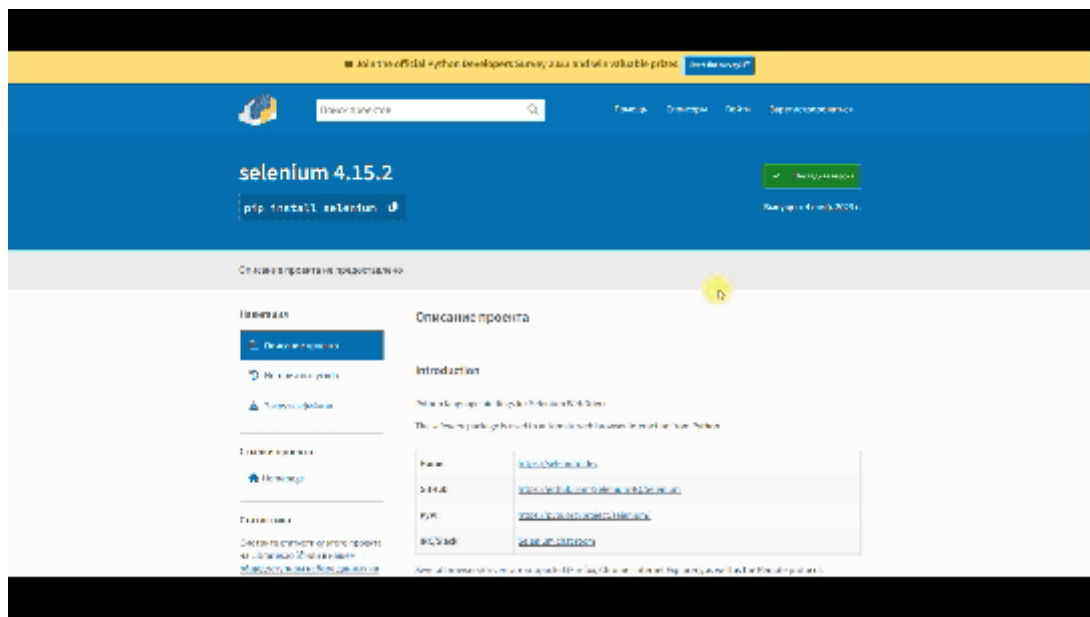
Для решения возьмем демо страницу, любезно предоставленную самой Рекапчей - <https://recaptcha-demo.appspot.com/recaptcha-v2-checkbox.php>:

Подготовка

На первом этапе необходимо все подготовить

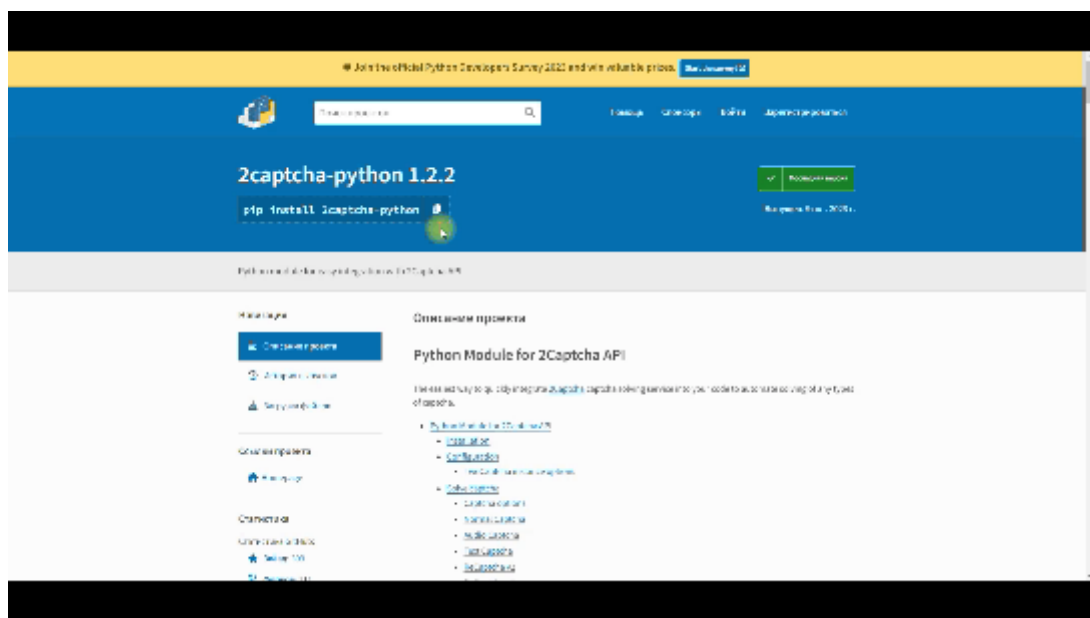
Для демонстрации работы я скачал следующие компоненты в специальную папку на своем компьютере (предварительно установил, конечно же Питон - но как я это сделал рассказывать не буду, надеюсь разберетесь).

Итак, нам понадобится - собственно сама библиотека для автоматизации браузера Селениум, берем тут - <https://pypi.org/project/selenium/>



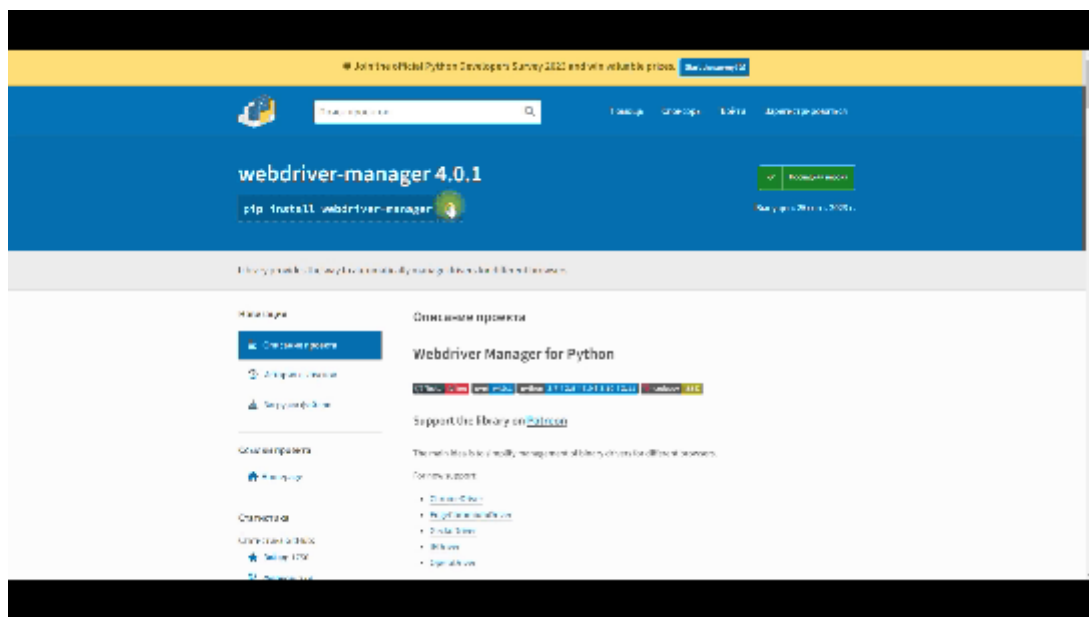
Установка Selenium

Официальный Python SDK для интеграции с API 2Captcha, берем тут <https://pypi.org/project/2captcha-python/>



Установка Python SDK 2captcha

SeleniumИ библиотека, которая упрощает загрузку и использование драйверов для Selenium, называется webdriver-manager, берем тут - <https://pypi.org/project/webdriver-manager/>



Установка webdriver-manager

Устанавливается все очень просто, копируете указанную команду и вставляете в консоли, как показано на видео выше

Можно устанавливать все по отдельности, как это сделал я а можете использовать универсальную команду

```
python -m pip install 2captcha-python selenium webdriver-manager
```

Поиск Sitekey

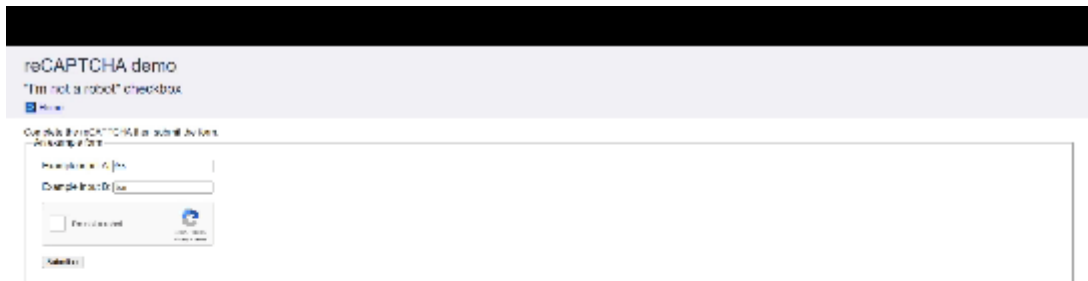
Так как речь идет про Рекапчу, то необходимо понимать, что такое параметр `site key`. Sitekey - это уникальный идентификатор, который Google присваивает всем своим формам с reCAPTCHA, и с помощью которого мы можем идентифицировать капчу на сайте.

Именно этот идентификатор и нужен нам, чтобы сервис распознавания капчи мог понять, с чем он имеет дело, и именно этот идентификатор мы и будем отправлять в 2капча.

Чтож, давайте искать sitekey на демо странице

- Переходим на <https://recaptcha-demo.appspot.com/recaptcha-v2-checkbox.php>
- Откройте инструменты разработки - нажать Ctrl/Cmd + Shift + I либо правую кнопку мыши - посмотреть код

- Найдите data-sitekey (нажмите ctrl + F, введите sitekey и нажмите enter) и скопируйте значение параметра.
- Сохраните значение, чтобы использовать его при отправке запроса для решения капчи на странице



Поиск SiteKey в код

Решение капчи

Естественно, для решения капчи нужно написать код - можете написать его самостоятельно, либо просто взять из уже готового мануала. Типа как из этого. Задача данного кода - попасть на целевую страницу и решить капчу через API. Именно то, чего мы и хотим достичь.

Сразу оговорюсь, в коде есть несколько параметров, которые необходимо изменить на свои, вот они сверху вниз:

2CAPTCHA_API_KEY - ваш АПИ, берете его в личном кабинете 2капча.

SITE_KEY - его мы с вами сохраняли на предыдущем шаге

Я очень сильно рассчитываю, что найти эти параметры в приведенном коде вы сможете самостоятельно

```
driver = webdriver.Chrome(service=ChromeService(ChromeDriverManager().install()))

captcha_page_url = "https://recaptcha-demo.appspot.com/recaptcha-v2-checkbox.php"
driver.get(captcha_page_url)
```

```
# решение капчи
```

```
print("solving captcha")
solver = TwoCaptcha("2CAPTCHA_API_KEY")
response = solver.recaptcha(sitekey='SITE_KEY', url=captcha_page_url)
code = response['code']
print(f"Successfully solved the Captcha. The solve code is {code}")
```

Этот кусок кода инициализирует объект `TwoCaptcha` с нашим ключом API и должен решать reCAPTCHA посредством вызова метода `recaptcha`. Для чего передаем значение `site key` и URL адрес страницы.

Это не полный код, читай дальше!

Отправка решенной капчи

Следующий кусок кода я не понял как вам правильно объяснить, так как сам не понял что происходит, поэтому просто перевел его и оставляю тут:

Далее находим элемент `g-recaptcha-response`, вставляем полученное значение для решения капчи и отправляем форму.

```
recaptcha_response_element = driver.find_element(By.ID, 'g-recaptcha-response')
driver.execute_script(f'arguments[0].value = "{code}";', recaptcha_response_element)

submit_btn = driver.find_element(By.CSS_SELECTOR, 'button[type="submit"]')
submit_btn.click()

input("Press enter to continue")
driver.close()
```

Полный код для автоматического решения reCaptcha

Собственно вот что мы получаем в результате.

```
from selenium.webdriver.common.by import By
from twocaptcha import TwoCaptcha
from selenium.webdriver.chrome.service import Service as ChromeService
from webdriver_manager.chrome import ChromeDriverManager
from selenium import webdriver
```

```
# Instantiate the WebDriver
driver = webdriver.Chrome(service=ChromeService(ChromeDriverManager().install()))

# Load the target page
captcha_page_url = "https://recaptcha-demo.appspot.com/recaptcha-v2-checkbox.php"
driver.get(captcha_page_url)

# Solve the Captcha
print("Solving Captcha")
solver = TwoCaptcha("2CAPTCHA_API_KEY")
response = solver.recaptcha(sitekey='SITE_KEY', url=captcha_page_url)
code = response['code']
print(f"Successfully solved the Captcha. The solve code is {code}")

# Set the solved Captcha
recaptcha_response_element = driver.find_element(By.ID, 'g-recaptcha-response')
driver.execute_script(f'arguments[0].value = "{code}";', recaptcha_response_element)

# Submit the form
submit_btn = driver.find_element(By.CSS_SELECTOR, 'button[type="submit"]')
submit_btn.click()

# Pause the execution so you can see the screen after submission before closing the
input("Press enter to continue")
driver.close()
```

Для того, чтобы продемонстрировать работоспособность кода, создам текстовый файл, назову его script.py и запущу в консоли. Результат вы можете увидеть ниже

Демо решение Рекапчи через Селениум

Для чистоты эксперимента решим этим же способом рекапчу с демо страницы, расположенной на сайте 2капча.

Нам понадобится поменять урл в коде и sitekey, смотрим что получилось:

Решение рекапчи с демо страницы 2капчи

Таким образом мы видим, что код рабочий. Пользуйтесь, но сперва не забудьте поблагодарить лайком автора.

Теги: selenium, webdriver, captcha, капча, решение капчи, recaptcha

Хабы: Python

Редакторский дайджест

Присылаем лучшие статьи раз в месяц

Электронпочта



12

0

Карма Рейтинг

Alex Gerasimchuk @kentavr009

Marketing Specialist

Комментарии 15



 **olegtsss**
10 ноя в 14:56

Не хватает теории и практики применения: что за reCAPTCHA, где и кем она применяется. Зачем ее нужно решать - из академического интереса или для практических задач (тогда каких)?

0 Ответить



 **Kolonist**
10 ноя в 16:04

что за reCAPTCHA, где и кем она применяется

Серьезно, на Хабре это надо описывать?



+3

Ответить



olegtsss

10 ноя в 16:18



Согласен, можно почитать на wiki. Но можно и написать, короткая вводная часть не будет лишней. Но вот практика - наше все. На хабре полно тех специалистов, но у каждого свой фундамент, ландшафт, технологический стек и т.д.



0

Ответить



Re3DaY

10 ноя в 16:40



Тот, кто не знает, что такое рекапча, вряд ли зайдет на эту статью



-2

Ответить



olegtsss

10 ноя в 17:03



Не согласен. Зайдет, например, чтобы узнать новое. При условии, что он, конечно, человек и заинтересовался не знакомым термином.



0

Ответить



kentavr009

11 ноя в 16:19



Есть серьезные сомнения, что зайдет. Вот вам бы было интересно почитать про синхрофазотрон и как его применять на гидроэлектростанции? Термины знакомые, но желания узнать про это побольше нет абсолютно.



0

Ответить



olegtsss

11 ноя в 16:31



Мне да, хотя я не знаю, что такое синхрофазотрон, но знаком с гидроэлектростанцией. Из-за этого «первого» - плюс нелепость сочетания указанных обозначений. Статью посмотрели 2300 человек, что дает не нулевую вероятность в защиту моей стороны.



0

Ответить



Tempest23

10 ноя в 16:40

Проблема, на мой взгляд, не до конца описана. Зачем, собственно, решать капчу в автотесте? Бывает надо проверить реальную интеграцию, но как будто бы не очень часто это происходит и не много усилий требует. В остальных тестах DUT конфигурируется так, чтобы она не мешала.

Что касается решения, то судя по коду, капчу решает сторонний сервис. Это рабочее решение, но не шибко интересное.

0 Ответить



 **kentavr009**
11 ноя в 16:20



Ну как сказать - я же поставил уровень сложности простой, а значит для новичков данное решение вполне себе вариант. И кстати, продукты ведь тоже можно самостоятельно купить в магазине, но кто-то заказывает доставку.

0 Ответить




 **yugo_fx**
11 ноя в 01:35

Как правило, рекапча-солверы используют скраперы данных.

Иногда для скрейпа данных совсем не обязательно использовать веб. Некоторые серверы например используют токен решения капчи для api запросов. Поэтому зная sitekey можно заранее решить капчу через какой-либо провайдер типа 2Captcha, capsolver или capmonster и отправить «защищенный капчей» запрос без селениумов.

+1 Ответить



 **yugo_fx**
11 ноя в 01:41



Другой вопрос, что это статья ради статьи) Ни практического смысла, ни инсайтов, ни идей. Автор, в чем мысль?

+1 Ответить



 **kentavr009**
11 ноя в 16:21



Можно, но в данном случае описан способ через Селениум)))

0 Ответить



 **S0mbre**
12 ноя в 16:14

Блин... просто скрипт для решения капчи через платный сервис. И все?.. А я думал, ИИ, вся мощь нейросетей, хитрая математика...

0 Ответить



 **kentavr009**
12 ноя в 19:10



Ну блин, не настолько я еще прокачан)) Но кому то и это будет полезно, согласитесь?

 +1 Ответить



 **S0mbre**
12 ноя в 23:36



Согласен, что ж))


 +1 Ответить



Только полноправные пользователи могут оставлять комментарии. Войдите, пожалуйста.

Публикации

ЛУЧШИЕ ЗА СУТКИ ПОХОЖИЕ

 **Tirarex**
22 часа назад

Обзор поддельного блока питания с закосом под Apple

 Простой  10 мин  12K

Обзор

 +69

 26

 81

 **Exosphere**
23 часа назад


Зачем писать на Хабр, если каждый раз потеют ладони?

 Простой  9 мин  4.6K

 +62

 24

 29

 **DRoman0v**
12 часов назад

В Canon разработали 5-нм литограф, который стоит дешевле, чем система от ASML. Но есть нюанс

 4 мин  16K

 +42

 12

 32



Tzimie
23 часа назад

Парадокс Белла для релятивистского паровозика

💧 Средний ⌚ 6 мин 👁 7.6K

📈 +39

🔖 30

💬 146



rip_m
21 час назад

Что нового в .NET 8?

⌚ 4 мин 👁 9K

📈 +27

🔖 31

💬 11



AlexeyW100
4 часа назад

Дизайн-система Gravity UI: как легко построить свой интерфейс

🟢 Простой ⌚ 7 мин 👁 2.9K

📈 +23

🔖 46

💬 0



ru_vds
2 часа назад

Мобильные суперприложения выгодны корпорациям, но это кошмар для простых людей

🟢 Простой ⌚ 7 мин 👁 1.7K

Мнение

📈 +21

🔖 7

💬 8



aszubarev
3 часа назад

Особенности сбора метрик. Запуск приложения gunicorn-ом в режиме мультипроцессинга

💧 Средний ⌚ 20 мин 👁 462

Тutorial

📈 +21

🔖 9

💬 0



BiktorSergeev
19 часов назад

Машина времени: hi-tech-новинки и цены на них в журналах начала 2000-х

🕒 9 мин 👁 3.6K

📌 +18

📖 11

💬 5



artmaro
4 часа назад

Хайп на AI или путевые заметки с выставки Gitex в Дубае

🕒 14 мин 👁 705

📌 +17

📖 5

💬 0

Ищем интересный техноконтент в Fashion retail

Опрос

Показать еще

ИСТОРИИ



Недельный топ-7 статей из блогов компаний

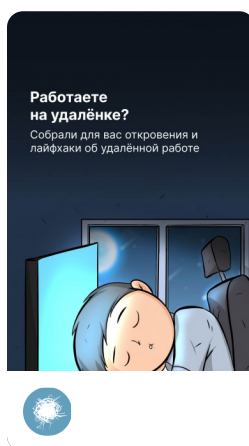
Топ-7 годных статей из блогов компаний



Сколько тратят в IT: сеньор бэкендер

Спрашиваем, сколько зарабатывают и тратят IT-специалисты по категориям бюджета: жилье, еда, хобби и прочие

Сколько тратят в IT: сеньор бэкендер



Работаете на удалёнке?
Собрали для вас откровения и лайфхаки об удалённой работе

Мифы и легенды удалёнки



Спасти бизнес за неделю до отпуска



Библиотека айтишника
Подборки новых и просто полезных книг для пополнения библиотеки.

Полезные книги для библиотеки айтишника



Пер кал доб соб

ВАКАНСИИ

Python разработчик

от 120 000 ₽ · Stroki · Можно удаленно

Python Backend

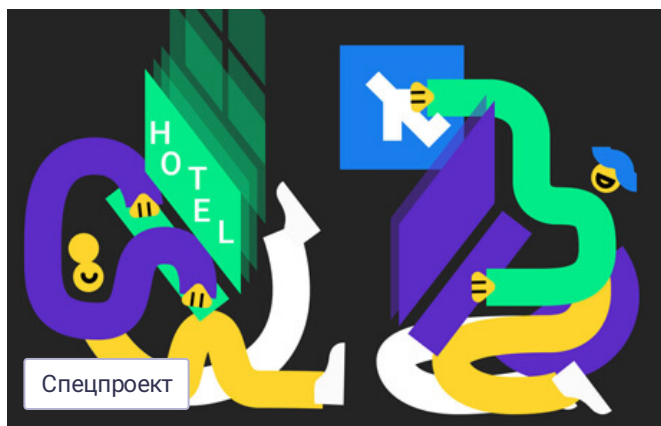
от 4 000 \$ · BullGlobal · Можно удаленно

Ведущий backend разработчик

от 130 000 ₽ · PromoData · Можно удаленно

Больше вакансий на Хабр Карьере

МИНУТОЧКУ ВНИМАНИЯ



Офер за неделю для бэкендеров



Где заканчивается мода и начинается IT?

РАБОТА

Django разработчик

49 вакансий

Python разработчик

134 вакансии

Data Scientist

87 вакансий

Все вакансии



 [Настройка языка](#)

[Техническая поддержка](#)

© 2006–2023, Habr