Отчет по НИРу

Примерная тема:

Автоматизирование прохождения CAPTCHA. Audio CAPTCHA

# Реферат

**Оглавление**

[Реферат 2](#__RefHeading___Toc102_1365281631)

[Введение 4](#__RefHeading___Toc104_1365281631)

[1. Современные методы защиты от ботов и спама на основе CAPTCHA 5](#__RefHeading___Toc122_1365281631)

[2. RECAPTCHA 5](#__RefHeading___Toc124_1365281631)

[3. HCAPTCHA 6](#__RefHeading___Toc126_1365281631)

[4. Capy 6](#__RefHeading___Toc153_1365281631)

[5. Автоматизированное тестирование С ИСПОЛЬЗОВАНИЕМ SELENIUM 7](#__RefHeading___Toc108_1365281631)

[6. Автоматизированное решение audio captcha как популярного метода защиты 7](#__RefHeading___Toc110_1365281631)

[Заключение 8](#__RefHeading___Toc112_1365281631)

[Список использованной литературы 9](#__RefHeading___Toc114_1365281631)

[Приложение 10](#__RefHeading___Toc116_1365281631)

# Введение

Коротко сказать про трудности автоматизации тестирования, когда есть капча

На сегодняшний день большинство web-ресурсов включают в себя различные меры по продиводействию спаммерам и ботам. Наиболее популярным и часто используемым является CAPTCHA в разных ее проявлениях.

# Современные методы защиты от ботов и спама на основе CAPTCHA

Проверочный код CAPTCHA – это одна из разновидностей мер безопасности, известная как аутентификация «вызов-ответ». Проверочный код защищает от спама и кражи паролей. Для проверки необходимо выполнить простой тест, подтверждающий, что действия выполняет человек, а не компьютерная программа, пытающаяся получить доступ к защищенному паролем аккаунту. https://support.google.com/a/answer/1217728?hl=ru

При проверке пользователь видит на экране два элемента: текстовое поле и случайную последовательность искаженных букв и/или цифр. Чтобы пройти проверку нужно ввести в поле изображенные символы.

На сегодняшний день наиболее часто встречаются следующие разновидности CAPTCHA:

1. reCAPTCHA
2. hCAPTCHA
3. Capy https://tproger.ru/articles/recaptcha-alternatives

# RECAPTCHA

<https://ru.wikipedia.org/wiki/ReCAPTCHA>

<https://en-m-wikipedia-org.translate.goog/wiki/ReCAPTCHA?_x_tr_sl=en&_x_tr_tl=ru&_x_tr_hl=ru&_x_tr_pto=sc>

reCAPTCHA — это система [CAPTCHA](https://en-m-wikipedia-org.translate.goog/wiki/CAPTCHA?_x_tr_sl=en&_x_tr_tl=ru&_x_tr_hl=ru&_x_tr_pto=sc) , принадлежащая [Google](https://en-m-wikipedia-org.translate.goog/wiki/Google?_x_tr_sl=en&_x_tr_tl=ru&_x_tr_hl=ru&_x_tr_pto=sc) . Она позволяет веб-хостерам различать человеческий и автоматизированный доступ к веб-сайтам. Первоначальная версия предлагала пользователям расшифровать трудночитаемый текст или сопоставить изображения. Версия 2 также предлагала пользователям расшифровать текст или сопоставить изображения, если анализ файлов cookie и визуализация холста предполагали, что страница загружается автоматически. Начиная с версии 3, reCAPTCHA никогда не будет прерывать работу пользователей и должна запускаться автоматически, когда пользователи загружают страницы или нажимают кнопки.

Отсканированный текст подвергается анализу двумя разными OCR. Любое слово, которое по-разному расшифровывается двумя программами OCR или отсутствует в английском словаре, помечается как «подозрительное» и преобразуется в CAPTCHA. Подозрительное слово отображается вне контекста, иногда вместе с уже известным контрольным словом. Если человек правильно вводит контрольное слово, то ответ на сомнительное слово принимается как вероятно действительный. Если достаточное количество пользователей правильно вводят контрольное слово, но неправильно вводят второе слово, которое OCR не распознал, то цифровая версия документов может содержать неправильное слово.

В 2012 году reCAPTCHA начала использовать фотографии, взятые из проекта [Google Street View](https://en-m-wikipedia-org.translate.goog/wiki/Google_Street_View?_x_tr_sl=en&_x_tr_tl=ru&_x_tr_hl=ru&_x_tr_pto=sc) , в дополнение к отсканированным словам. Пользователю будет предложено идентифицировать изображения пешеходных переходов, уличных фонарей и других объектов.

reCAPTCHA v1 была объявлена [​​устаревшей](https://en-m-wikipedia-org.translate.goog/wiki/End-of-life_product?_x_tr_sl=en&_x_tr_tl=ru&_x_tr_hl=ru&_x_tr_pto=sc) и закрыта 31 марта 2018 года.

В 2013 году reCAPTCHA начала внедрять поведенческий анализ взаимодействий браузера, чтобы предсказать, является ли пользователь человеком или ботом. В следующем году Google начала развертывать новый API reCAPTCHA, включающий «no CAPTCHA reCAPTCHA» — где пользователи, считающиеся имеющими низкий риск, должны только нажать на один [флажок](https://en-m-wikipedia-org.translate.goog/wiki/Checkbox?_x_tr_sl=en&_x_tr_tl=ru&_x_tr_hl=ru&_x_tr_pto=sc) , чтобы подтвердить свою личность. CAPTCHA все еще может быть представлена, если система не уверена в риске пользователя; Google также представила новый тип CAPTCHA-задачи, разработанный для того, чтобы быть более доступным для мобильных пользователей, где пользователь должен выбрать изображения, соответствующие определенному запросу из сетки.

В 2017 году Google представила новую «невидимую» reCAPTCHA, где проверка происходит в фоновом режиме, и никакие проблемы не отображаются вообще, если пользователь считается малорискованным.

# HCAPTCHA

# Capy

# Автоматизированное тестирование С ИСПОЛЬЗОВАНИЕМ SELENIUM

Здесь короткая теория про Selenium

<https://ru.wikipedia.org/wiki/Selenium>

Selenium WebDriver — инструмент для автоматизации действий веб-браузера. В большинстве случаев используется для тестирования [Web-приложений](https://ru.wikipedia.org/wiki/Веб-приложение), но этим не ограничивается. В частности, он может быть использован для решения рутинных задач администрирования сайта или регулярного получения данных из различных источников (сайтов).

# Автоматизированное решение audio captcha как популярного метода защиты

Здесь будет описана собственно практическая часть

Для разработки сервиса автоматизации распознавания CAPTCHA был выбран язык программирования Python и библиотека для автоматизации тестирования web-приложений Selenium.

Audio CAPTCHA представляет собой элемент, встроенный в web-страницу, который содержит в себе ссылку на отрезок звуковой дорожки, которая содержит шум и запись голоса.

Данная запись хорошо поддается распознаванию с использованием библиотек для распознавания речи, одна из которых была использована для обхода Audio CAPTCHA в данной работе.

В Python есть ряд библиотек для распознавания человеческой речи, которые предоставляют доступ к API от нескольких компаний среди которых: ПЕРЕЧИСЛИТЬ БИБЛИОТЕКИ.

Библиотека SpeechRecognition была выбрана по нескольким причинам: ПЕРЕЧИСЛИТЬ ПРИЧИНЫ.

Тут написать про алгоритм распознавателя, а затем про алгоритм автоматизации

# Заключение

Дать резюме о проделанной работе

# Список использованной литературы

# Приложение