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 метод защиты, основанный на принципе аутентификации «вызов-ответ». Он предназначен для предотвращения автоматических действий, таких как спам или попытки взлома учетных записей, путем выполнения пользователем простого теста, подтверждающего, что он человек, а не программа [1].

В процессе проверки на экране отображаются два элемента: текстовое поле и случайная последовательность символов, таких как искаженные буквы, цифры или другие изображения. Пользователю необходимо ввести эти символы в поле ввода, чтобы доказать свою подлинность.

CAPTCHA является важной мерой безопасности, так как предотвращает автоматические атаки, например, массовую регистрацию ботов, и защищает данные пользователя. Современные системы CAPTCHA используют не только текст, но и изображения, аудио, поведенческие анализы и другие инновационные подходы, чтобы сделать тесты удобными для людей, но сложными для программ.

На сегодняшний день наиболее распространенные виды CAPTCHA включают:

1. reCAPTCHA – разработанная Google система, которая предлагает тесты на основе распознавания объектов, анализа поведения или текстовых символов.
2. hCAPTCHA – альтернатива reCAPTCHA, фокусирующаяся на защите конфиденциальности пользователей.
3. Capy – система CAPTCHA, предлагающая пользователю головоломки, например, сборку изображения или взаимодействие с элементами интерфейса [2].

## reCAPTCHA

reCAPTCHA – система защиты от автоматизированных действий, разработанная Google, которая помогает различать человека и бота. Она объединяет несколько подходов, делая проверку удобной для пользователей, но сложной для автоматических систем [3].

reCAPTCHA включает в себя следующие версии:

1. reCAPTCHA v1 (устарела в 2018 году):
   1. Пользователи вводили текст, состоящий из искаженных слов, отображаемых на изображении.
   2. Использовала слова из книг и документов, которые не могли быть распознаны OCR.
2. reCAPTCHA v2:
   1. Клик по флажку: пользователи подтверждают, что они не роботы, нажимая на флажок «Я не робот».
   2. Выбор объектов на изображениях: пользователи идентифицируют заданные объекты на сетке из картинок.
   3. Аудио CAPTCHA: для пользователей с ограничениями зрения, предлагается прослушать запись и ввести услышанные символы.
3. reCAPTCHA v3:
   1. Полностью работает в фоновом режиме, анализируя поведение пользователя на странице.
   2. Не требует явных действий, если пользователь считается низкорискованным [4].

## hCAPTCHA

hCAPTCHA – это альтернативная система CAPTCHA, разработанная для защиты сайтов от ботов и спама, при этом уделяющая особое внимание конфиденциальности пользователей. Она стала популярной благодаря своей гибкости и ориентации на защиту данных [5].

Основные особенности hCAPTCHA:

1. Конфиденциальность:
   1. В отличие от reCAPTCHA, hCAPTCHA не собирает данные о пользователях для рекламных целей, что делает ее привлекательной с точки зрения соблюдения конфиденциальности.
2. Простота интеграции:
   1. Легко интегрируется с веб-сайтами через API.
   2. Совместима с большинством популярных платформ, таких как WordPress, и может быть настроена для разных типов взаимодействия.
3. Модели монетизации:
   1. Владельцы сайтов могут зарабатывать, разрешая hCAPTCHA использовать проверочные задачи, связанные с машинным обучением, например, разметку данных.

Виды взаимодействия с пользователями:

1. Графическая CAPTCHA: выбор изображений, соответствующих запросу.
2. Текстовая CAPTCHA: ввод символов (редко используется).
3. Аудио CAPTCHA: для пользователей с ограниченными возможностями, предлагается прослушать и ввести услышанные символы.
4. Клик CAPTCHA: нажатие на флажок «Я не робот» (для низкорискованных пользователей).

## Capy

Capy CAPTCHA – это инновационная система CAPTCHA, разработанная с акцентом на удобство для пользователей и адаптацию к современным веб-средам. Она предлагает интерактивные методы проверки, направленные на минимизацию раздражения пользователей при сохранении высокого уровня защиты от ботов.

Основные особенности Capy CAPTCHA:

1. Интерактивность:
   1. Capy использует методы проверки, которые требуют не просто ввода текста или выбора картинок, а выполнения задач, таких как перемещение объектов.
   2. Простые задачи делают процесс проверки менее раздражающим и более интуитивным.
2. Гибкость настройки:
   1. Система может быть адаптирована под конкретные нужды сайта, включая выбор сложности задач и дизайн интерфейса.
3. Доступность:
   1. Подходит для пользователей с различными потребностями, включая мобильные устройства.

Виды взаимодействия с пользователями:

1. Головоломки (Puzzle CAPTCHA): сборка пазла с перемещением недостающих элементов в нужное место.
2. Тесты на логику и распознавание: выбор нужного объекта или логического варианта из предложенных.
3. Текстовая CAPTCHA (редко используется).

Capy CAPTCHA используется на сайтах, где важны как защита от ботов, так и положительный пользовательский опыт. Особенно популярна в проектах с высоким акцентом на дизайн и пользовательское взаимодействие.

# 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

1. <https://support.google.com/a/answer/1217728?hl=ru>
2. <https://tproger.ru/articles/recaptcha-alternatives>
3. <https://ru.wikipedia.org/wiki/ReCAPTCHA>
4. <https://en.m.wikipedia.org/wiki/ReCAPTCHA>
5. <https://www.hcaptcha.com/>

# Приложение