**«Санкт-Петербургский государственный электротехнический университет**

**«ЛЭТИ» им. В.И. Ульянова (Ленина)»**

**(СПбГЭТУ «ЛЭТИ»)**

|  |  |  |
| --- | --- | --- |
| **Направление** | [09.03.04 Программная инженерия](http://www.eltech.ru/ru/abiturientam/napravleniya-podgotovki/bakalavriat/01040062-prikladnaya-matematika-i-informatika) | |
| **Профиль** | Разработка программно-информационных систем | |
| **Факультет** | КТИ | |
| **Кафедра** | МО ЭВМ | |
| *К защите допустить* |  | |
| И.о. зав. кафедрой |  | А.А. Лисс |

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА**

**БАКАЛАВРА**

**Тема:** **РАЗРАБОТКА СРЕДСТВА** **ВИЗУАЛИЗАЦИИ И МОНИТОРИНГА КРИМИНАЛЬНОГО КОНТЕНТА ИЗ ТУМАННОЙ ВЫЧИСЛИТЕЛЬНОЙ СРЕДЫ**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Студент |  |  | |  | Л.С. Рослова |
|  |  | *подпись* | |  |  |
| Руководитель | к.т.н., доцент |  | |  | А.А. Лисс |
|  |  | *подпись* | |  |  |
| Консультанты | к.т.н., доцент |  | |  | А.Ю. Первицкий |
|  |  | *подпись* | |  |  |
|  |  |  | |  | А.Н. Субботин |
|  |  | *подпись* | |  |  |
|  | к.т.н. |  | |  | М.М. Заславский |
|  |  | *подпись* | |  |  |
|  |  | |  |  |  |

Санкт-Петербург

2023

**ЗАДАНИЕ**

**НА ВЫПУСКНУЮ КВАЛИФИКАЦИОННУЮ РАБОТУ**

|  |  |
| --- | --- |
|  | Утверждаю |
|  | И.о. зав. кафедрой МО ЭВМ |
|  | \_\_\_\_\_\_\_\_\_\_\_\_ А.А. Лисс |
|  | «\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_2023 г. |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Студент | Рослова Л.С. | | |  | Группа | | 9304 | |
| Тема работы: Разработка средства визуализации и мониторинга криминального контента из туманной вычислительной среды | | | | | | | | |
| Место выполнения ВКР: Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина) | | | | | | | | |
| Исходные данные (технические требования):  ОС Windows, ОС Linux | | | | | | | | |
| Содержание ВКР:  Введение, Обзор предметной области, Функциональные требования,  Архитектура системы, Исследование разработанного решения,  Доп.раздел БЖД, Заключение | | | | | | | | |
| Перечень отчетных материалов: пояснительная записка, иллюстративный материал. | | | | | | | | |
| Дополнительные разделы: Безопасность жизнедеятельности. | | | | | | | | |
|  | | | | | | | | |
| Дата выдачи задания | | | Дата представления ВКР к защите | | | | | |
| «\_04\_»\_\_апреля\_\_\_\_\_2023 г. | | | «9»\_\_июня\_\_\_2023 г. | | | | | |
|  | | |  | | | | | |
| Студент | |  | | | | Л.С. Рослова | |  |
| Руководитель к.т.н., доцент | |  | | | | А.Ю. Первицкий | |  |
| Консультант | |  | | | | А.Н. Субботин | |  |

**КАЛЕНДАРНЫЙ ПЛАН ВЫПОЛНЕНИЯ**

**ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ**

|  |  |
| --- | --- |
|  | Утверждаю |
|  | И.о. зав. кафедрой МО ЭВМ |
|  | \_\_\_\_\_\_\_\_\_\_\_\_ А.А. Лисс |
|  | «\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_2021 г. |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Студент | Рослова Л.С. |  | Группа | 9304 |
| Тема работы: Разработка средства визуализации и мониторинга криминального контента из туманной вычислительной среды | | | | |

|  |  |  |
| --- | --- | --- |
| № п/п | Наименование работ | Срок выполнения |
| 1 | Обзор литературы по теме работы | 04.04 – 03.05 |
| 2 | Обзор предметной области | 04.05 – 06.05 |
| 3 | Постановка задачи и формализация функциональных требований | 06.05 – 10.05 |
| 4 | Архитектура программной реализации | 10.05 – 12.05 |
| 5 | Результаты разработки/исследование разраб.системы | 12.05 – 22.05 |
| 6 | БЖД | 22.05 – 23.05 |
| 7 | Оформление пояснительной записки | 23.05 – 31.05 |
| 8 | Оформление иллюстративного материала | 01.06 – 04.06 |
| 9 | Предзащита | 01.06 |

|  |  |  |  |
| --- | --- | --- | --- |
| Студент | |  | Рослова Л.С. |
| Руководитель к.т.н., доцент | |  | А.Ю. Первицкий |
| Консультант | |  | А.Н. Субботин |
|  |  |  |  |

**РЕФЕРАТ**

Пояснительная записка 65 стр., 19 рис., 16 табл., 24 ист.

КРИМИНАЛЬНЫЙ КОНТЕНТ, МОНИТОРИНГ, АНАЛИЗ, ИНТЕРНЕТ-РЕСУРС, МОДЕЛЬ, НЕЙРОННАЯ СЕТЬ, ТУМАННЫЕ ТЕХНОЛОГИИ

**Объектом исследования** являются интернет-ресурсы, содержащие криминальный контент.

**Предметом исследования** являются методы анализа контента веб-сайтов с использованием машинного обучения.

**Цель работы:** разработка архитектуры расширения для браузера анализирующего посещаемый веб-ресурс на криминальный контент с применением средств машинного обучения и с возможностью минимизации задержек при обработке данных.

Результатом данной работы является создание расширение для браузера, которое анализирует изображения на интернет-ресурсе моделями глубокого обучения и предоставляет информацию о том, к какой категории оно относится. Для этого использовалась готовые программные библиотеки, позволяющие строить и обучать нейронной сети с целью автоматической классификации образов. В работе описывается архитектура расширения, состоящая из двух основных компонентов: модуля загрузки изображений и отображения результатов анализа, модуля загрузки и подключения модели к расширению.

**ABSTRACT**

The result of this work is the creation of a browser extension that analyzes images on an Internet resource with deep learning models and provides information about which category it belongs to. To do this, we used ready-made software libraries that allow us to build and train neural networks in order to automatically classify images. The paper describes the architecture of the extension, consisting of two main components: the module for loading images and displaying analysis results, the module for loading and connecting the model to the extension.

**Содержание**

[ОПРЕДЕЛЕНИЯ, ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ 7](#__RefHeading___Toc1844_2415813192)

[ВВЕДЕНИЕ 8](#__RefHeading___Toc1846_2415813192)

[​ 1 ОБЗОР ПРЕДМЕТНОЙ ОБЛАСТИ 11](#__RefHeading___Toc1848_2415813192)

[​ 1.1 Обработка и анализ изображений 11](#__RefHeading___Toc1850_2415813192)

[​ 1.1.1 Возможности обработки изображений моделями машинного обучения 11](#__RefHeading___Toc1970_2415813192)

[1.1.2 Нейронные сети для классификации изображений 11](#__RefHeading___Toc1972_2415813192)

[1.1.2.1 VGG-16 11](#__RefHeading___Toc1974_2415813192)

[1.1.2.2 Inception 11](#__RefHeading___Toc1976_2415813192)

[1.1.2.3 MobileNetV2 11](#__RefHeading___Toc1978_2415813192)

[1.1.2.4 ResNet50 11](#__RefHeading___Toc1980_2415813192)

[1.1.2.5 EfficientNet 11](#__RefHeading___Toc1982_2415813192)

[​ 1.2 Анализ существующих систем мониторинга криминального контента 11](#__RefHeading___Toc1984_2415813192)

[​ 1.3 Выводы 11](#__RefHeading___Toc1986_2415813192)

[ЗАКЛЮЧЕНИЕ (заголовок второго уровня) 12](#__RefHeading___Toc1854_2415813192)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ (заголовок второго уровня) 13](#__RefHeading___Toc1856_2415813192)

[ПРИЛОЖЕНИЕ А (заголовок второго уровня) 14](#__RefHeading___Toc1858_2415813192)

**ОПРЕДЕЛЕНИЯ, ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ**

ЯП – язык программирования;

HTML – стандартизированный язык разметки документов для просморта веб-страниц в браузере;

Датасет – набор данных для обучения нейронной сети;

JSON – текстовый формат обмена данными;

Тензор – обобщение векторов и матриц на более высокие измерения.

## ВВЕДЕНИЕ

Согласно данным из отчёта ‘Global Digital 2022’[1] процент интернет-пользователей в 2022 году перешёл черту в 62% от населения мира. При этом каждый имеет возможность выкладывать в сеть новый материал. Особенно активно в сеть загружаются изображения, которые практически никогда не проходят проверку на криминальное содержимое перед публикацией. Актуальность расширения для браузера, анализирующего такой материал на криминальное содержание с помощью нейронной сети, обусловлена необходимостью защиты пользователей от нежелательного контента в интернете. В современном мире существует большое количество сайтов, на которых имеется деструктивное содержание, такое как порнография, алкоголь, наркотики, оружие и т.д. Этот контент может быть вредным для психического и эмоционального здоровья пользователей, особенно детей и подростков.

Традиционные методы блокировки нежелательного контента, такие как фильтры URL и списки блокировки сайтов, не всегда эффективны и могут пропустить определенные типы деструктивного материала. Однако использование методов машинного обучения, в том числе нейронных сетей, позволяет более точно и надежно определять деструктивное содержание на изображениях.

Таким образом, разработка расширения для браузера, анализирующего изображения на криминальный контент с помощью нейронной сети, является актуальной задачей, которая может улучшить безопасность пользователей в интернете и предотвратить негативные последствия, связанные с просмотром нежелательного материала.

**Цель работы:** разработка архитектуры расширения для браузера анализирующего посещаемый веб-ресурс на криминальный контент с применением средств машинного обучения и с возможностью минимизации задержек при обработке данных.

**Задачи данной работы:**

1. Изучение методов и технологий машинного обучения, необходимых для анализа изображений.
2. Разработка архитектуры расширения для браузера, которое будет осуществлять анализ посещаемых веб-ресурсов на наличие криминального контента.
3. Реализация механизмов, обеспечивающих сбор и анализ данных о посещаемых веб-ресурсах.
4. Интеграция средств машинного обучения для обработки и анализа данных на наличие криминального контента.
5. Тестирование и оптимизация работы расширения для браузера, с целью повышения точности определения криминального контента и уменьшения ложных срабатываний.

**Объектом исследования** являются интернет-ресурсы, содержащие криминальный контент.

**Предметом исследования** являются методы анализа контента веб-сайтов с использованием машинного обучения.

**Практическая ценность работы:**

1. Повышение безопасности пользователей в интернете. Расширение может обнаруживать наличие криминального контента на посещаемых веб-ресурсах, что поможет избежать неприятных ситуаций и предотвратить потенциальные угрозы для пользователей.
2. Уменьшение распространения криминального контента в интернете. Расширение может обнаруживать криминальный контент на посещаемых веб-ресурсах и предупреждать пользователей о его наличии. Это может помочь снизить распространение криминального контента в интернете и сделать его менее доступным.
3. Упрощение родительского контроля. Родители могут использовать расширение для браузера, чтобы отслеживать посещаемые их детьми веб-ресурсы и обнаруживать наличие криминального контента на этих ресурсах.

## 1 **ОБЗОР ПРЕДМЕТНОЙ ОБЛАСТИ**

### 1.1 Обработка и анализ изображений

#### 1.1.1 **Возможности обработки изображений моделями машинного обучения**

#### **1.1.2 Нейронные сети для классификации изображений**

#### **1.1.2.1 VGG-16**

#### **1.1.2.2 Inception**

#### **1.1.2.3 MobileNetV2**

#### **1.1.2.4 ResNet50**

#### **1.1.2.5 EfficientNet**

### 1.2 **Анализ существующих систем мониторинга криминального контента**

### **1.3 Выводы**

Пример основного текста



Рисунок 1 – Пример рисунка

Каждая таблица должна иметь ссылку в тексте (таблица 1).

Таблица 1 – Пример таблицы

|  |  |  |
| --- | --- | --- |
| Показатель | Единица измерения | Значение |
| Время работы | сек. | 159 |
| Нагрузка на ЦПУ | процент | 95 |
| Стоимость работ | тыс. руб. | 100 |

## ЗАКЛЮЧЕНИЕ (заголовок второго уровня)

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ (заголовок второго уровня)

1. Обзор Digital 2022 [Электронный ресурс] URL: https://www.web-canape.ru/business/statistika-interneta-i-socsetej-na-2022-god-cifry-i-trendy-v-mire-i-v-rossii/ (дата обращения: 23.12.2022)
2. Бобкова, О.В., Давыдов, С.А., Ковалева, И.А., Плагиат как гражданское правонарушение / О.В. Бобкова, С.А. Давыдов, И.А. Ковалева // Патенты и лицензии. – 2016. – № 7. – С. 31-37.

## ПРИЛОЖЕНИЕ А (заголовок второго уровня)