

Федеральное государственное автономное
образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Космических и информационных технологий
институт
Высокопроизводительные вычисления
кафедра

УТВЕРЖДАЮ
Заведующий кафедрой
_____ Д.А. Кузьмин
подпись инициалы, фамилия
«_____» _____ 2020 г.

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Библиотека для оптимизации работы с изображениями в высоконагруженных
Web-приложениях
тема

09.04.01 Информатика и вычислительная техника
код и наименование направления

09.04.01.01 «Высокопроизводительные вычислительные системы»
код и наименование магистерской программы

Научный руководитель	_____	<u>зав. каф. ВПВ</u> <u>канд.техн.наук, доцент</u> должность, ученая степень	<u>Д.А. Кузьмин</u> инициалы, фамилия
Выпускник	_____		<u>Н.Э. Кирилов</u> инициалы, фамилия
Рецензент	_____	<u>канд.техн.наук,</u> <u>генеральный директор</u> <u>ООО «Инфодор-Инфо»</u> должность, ученая степень	<u>Э.А. Цокур</u> инициалы, фамилия
Нормоконтролер	_____	<u>зав. каф. ВПВ</u> <u>канд.техн.наук., доцент</u> должность, ученая степень	<u>Д.А. Кузьмин</u> инициалы, фамилия

Красноярск 2020

Федеральное государственное автономное
образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Институт космических и информационных технологий

институт

Высокопроизводительные вычисления

кафедра

УТВЕРЖДАЮ
Заведующий кафедрой

_____ Д.А. Кузьмин
подпись инициалы, фамилия
« _____ » _____ 2020г.

**ЗАДАНИЕ
НА ВЫПУСКНУЮ КВАЛИФИКАЦИОННУЮ РАБОТУ
в форме магистерской диссертации**

Студенту Кирилову Никите Эдуардовичу

Группа КИ18-01-1М. Направление (специальность) 09.04.01, «Информатика и вычислительная техника».

Тема выпускной квалификационной работы: «Библиотека для оптимизации работы с изображениями в высоконагруженных Web-приложениях».

Утверждена приказом по университету № от _____
Руководитель ВКР Д.А. Кузьмин, канд. техн. наук, доцент, заведующий кафедрой «Высокопроизводительных вычислений».

Перечень разделов ВКР: Введение, анализ технического задания, разработка основных технических решений, руководство пользователя, заключение.

Перечень графического или иллюстративного материала с указанием основных чертежей, плакатов, слайдов: презентационные слайды PDF.

Руководитель ВКР

подпись, дата

Д.А. Кузьмин
инициалы, фамилия

Задание принял к исполнению

подпись, дата

Н.Э. Кирилов
инициалы, фамилия

« ____ » _____ 2020 г.

РЕФЕРАТ

Выпускная квалификационная работа по теме «Библиотека для оптимизации работы с изображениями в высоконагруженных Web-приложениях» содержит 53 страницы, 17 рисунков, 3 таблицы, 12 использованных источников и 2 приложения.

БИБЛИОТЕКА, ФОРМАТЫ ИЗОБРАЖЕНИЙ, ОПТИМИЗАЦИЯ, PYTHON, БИНАРНЫЕ КЛАССИФИКАТОРЫ, МАШИННОЕ ОБУЧЕНИЕ

Цель работы: разработка программной библиотеки для оптимизации работы с JPEG-изображениями, используемых в веб-приложениях.

В данной работе были сформулированы требования к библиотеке, а также проанализированы средства для ее реализации, а именно: сбор данных для обучения бинарного классификатора, анализ данных, обучение классификатора.

В результате анализа требований была разработана общая архитектура библиотеки на уровне модулей. Затем был осуществлен процесс написания программного кода для разработанной архитектуры и сформулированных требований к функционалу библиотеки.

2.4.2.3 Анализ изображений преимущественного синего цвета.....	26
2.4.2.4 Общий вывод по анализу цветов.....	26
2.4.3 Анализ набора данных изображений при помощи программных средств.....	28
2.4.3.1 Визуализация основных характеристик из набора даннх.....	28
2.4.3.2 Корреляция основных характеристик в наборе данных.....	30
2.4.4 Допустимые классы при работе классификатора.....	32
2.4.5 Выводы по анализу.....	32
2.5 Общая архитектура библиотеки.....	33
2.6 Модуль определения параметров изображений.....	34
2.7 Модуль классификации.....	34
2.7.1 Классификаторы.....	36
2.7.2 Выбор алгоритма классификации.....	38
2.7.3 Сохранение и последующее использование модели классификатора.	40
2.8 Вывод по разделу.....	41
3 Руководство пользователя.....	43
3.1 Установка и запуск библиотеки.....	43
3.2 Примеры использования библиотеки.....	43
3.2.1 Оптимизация объема хранимых изображений.....	43
3.2.2 Оптимизация веса WEB-страницы.....	44
3.3 Тестирование программы под разными операционными системами.....	45
ЗАКЛЮЧЕНИЕ.....	46
СПИСОК СОКРАЩЕНИЙ.....	47
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ.....	48
ПРИЛОЖЕНИЕ А.....	49
ПРИЛОЖЕНИЕ Б.....	52

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	4
1 Анализ технического задания.....	5
1.1 Цель и задачи проектирования.....	5
1.2 Анализ необходимости оптимизации работы с изображениями в веб-приложениях.....	5
1.3 Обзор существующих способов для оптимизации работы с изображениями.....	10
1.3.1 Изменение размеров изображений.....	10
1.3.2 Сжатие изображений.....	11
1.3.3 Использование Content Delivery Network (CDN).....	12
1.3.4 Перспективные форматы изображений.....	12
1.3.4.1 Формат WEBP.....	13
1.3.4.2 Формат AVIF.....	13
1.4 Техническое задание.....	14
1.4.1 Наименование системы.....	14
1.4.2 Назначение и цели создания библиотеки.....	14
1.4.3 Требования к библиотеке.....	15
1.4.3.1 Требования к структуре и функциональной части библиотеки.....	15
1.4.3.2 Требование к средствам информационного обмена между компонентами системы.....	15
1.4.3.3 Требования к видам обеспечения.....	16
1.5 Перспективы развития программы проверки.....	18
1.6 Выводы по разделу.....	18
2 Разработка основных технических решений.....	19
2.1 Общий принцип работы библиотеки.....	19
2.2 Определение входных и выходных данных.....	20
2.3 Сбор набора данных с изображениями.....	20
2.4 Анализ набора изображений для обучения классификатора.....	22
2.4.1 Анализ на основании веса изображений.....	22
2.4.2 Анализ влияния базового цвета изображения на его вес после преобразования.....	23
2.4.2.1 Анализ изображений преимущественного красного цвета.....	24
2.4.2.2 Анализ изображений преимущественного зеленого цвета.....	25

ВВЕДЕНИЕ

На сегодняшний день Интернет занимает все более и более значимую роль в нашей жизни. Это обусловлено тем, что число людей, у которых есть доступ в Интернет, с каждым годом постоянно растет. В апреле 2020 число активных пользователей превысило 4.5 миллиарда и это число, безусловно, не предел [1].

Все большая доступность Интернета приводит также и к тому, что на сегодняшний день существует огромное число веб-сайтов различной направленности и сложности реализации (на текущий момент – 1.7 миллиарда, что почти в 10 раз больше, чем 10 лет назад). Все это привело к тому, что время, в котором мы живем по праву можно назвать «золотым веком» веб-приложений. Во многом это обусловлено не только сильно возросшим числом активных пользователей, но и тем, как сильно Интернет позволяет упростить и автоматизировать различные процессы, а также упростить коммуникацию между людьми и организациями и обмен данными между ними.

Считается, что хорошим является то веб-приложение, которое, в первую очередь, удобно для использования конечным пользователем. Удобство работы с приложением можно рассматривать по некоторым критериям. К таким критериям можно отнести удобный, визуально красивый и интуитивно понятный графически интерфейс. Так же удобством можно считать широкий функционал, общая стабильность, безопасность и надежность работы. Однако, одним из решающих критериев удобства все-таки считается быстрая скорость работы. Именно поэтому задача по улучшению скорости работы веб-приложений постоянно является одной из самых актуальных для программистов в процессе разработки и поддержки.

1 Анализ технического задания

1.1 Цель и задачи проектирования

Целью настоящей выпускной квалификационной работы является разработка программной библиотеки для оптимизации работы с JPEG-изображениями, используемых в веб-приложениях.

Для реализации поставленной цели необходимо решение следующих задач:

- 1) Проанализировать необходимость оптимизации работы с изображениями в веб-приложениях;
- 2) Выполнить обзор существующих способов для оптимизации работы с изображениями в веб-приложениях;
- 3) Разработать техническое задание на основании проведенного анализа аналогов;
- 4) Выбрать инструментальные средства для написания библиотеки;
- 5) Выполнить процесс проектирования;
- 6) Реализовать техническое задание;
- 7) Произвести тестирование библиотеки;
- 8) Написать руководство пользователя и программиста;
- 9) Рассмотреть способ интеграции библиотеки в архитектуру существующего веб-приложения.

1.2 Анализ необходимости оптимизации работы с изображениями в веб-приложениях

На рисунке 1 представлен график того, как за последние 10 лет менялось значение времени загрузки страницы на персональных компьютерах (ПК) и мобильных устройствах (МУ) – одной из главных характеристик скорости работы веб приложения. Согласно этому графику, скорость полной загрузки веб-страницы составляет при медианном значении 6.5с и 19.6с для

Изъяты страницы с 6 по 45

ЗАКЛЮЧЕНИЕ

В результате выполнения выпускной квалификационной работы магистра была разработана библиотека на языке Python, которую можно использовать для оптимизации работы с JPEG-изображениями, используемых в высоконагруженных Web-приложениях.

СПИСОК СОКРАЩЕНИЙ

ПК – персональный компьютер

МУ – мобильное устройство


СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Minnesota Internet Traffic Studies (MINTS) [Электронный ресурс] : – Режим доступа: <http://www.dtc.umn.edu/mints/home.php>
2. Cisco Annual Internet Report (2018–2023) [Электронный ресурс] : – Режим доступа: <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.pdf>
3. Image Optimization [Электронный ресурс] : – Режим доступа: <https://developers.google.com/web/fundamentals/performance/optimizing-content-efficiency/image-optimization>
4. Robinson, D. Content Delivery Networks: Fundamentals, Design, and Evolution :печатное издание / Don Robinson. – Wiley, 2017. – 256.
5. A new image format for the Web [Электронный ресурс] : – Режим доступа: <https://developers.google.com/speed/webp/>
6. AV1 Image File Format (AVIF) [Электронный ресурс] : – Режим доступа: <https://aomediacodec.github.io/av1-avif/>
7. Unsplash Developers [Электронный ресурс] : – Режим доступа: <https://unsplash.com/developers>
8. Geron, A. Hands-On Machine Learning with Scikit-Learn and TensorFlow :печатное издание / Aurélien Géron. – O’Reilly, 2017. – 566 с.
9. Raschka, S. Python Machine Learning :печатное издание / Sebastian Raschka. – Packt Publishing, 2015. – 456 с.
10. Grus, J. Data Science from scratch :печатное издание / Joel Grus. – O’Reilly, 2019. – 500с.
11. Chollet, F. Data Science from scratch :печатное издание / Francois Chollet. – Manning Publications, 2017. – 384с.
12. Buduma, N. Fundamentals of deep learning :печатное издание / Nikhill Buduma. – O’Reilly, 2015. – 150.

Изъяты страницы с 49 по 53

Федеральное государственное автономное
образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Космических и информационных технологий
институт
Высокопроизводительные вычисления
кафедра

УТВЕРЖДАЮ
Заведующий кафедрой
 Д.А. Кузьмин
подпись инициалы, фамилия
« 17 » 06 2020 г.

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Библиотека для оптимизации работы с изображениями в
высоконагруженных Web-приложениях
тема

09.04.01 Информатика и вычислительная техника
код и наименование направления

09.04.01.01 «Высокопроизводительные вычислительные системы»
код и наименование магистерской программы

Научный руководитель	 подпись, дата	зав. каф. ВПВ канд.техн.наук, доцент должность, ученая степень	Д.А. Кузьмин инициалы, фамилия
Выпускник	 подпись, дата		Н.Э. Кирилов инициалы, фамилия
Рецензент	 подпись, дата	канд.техн.наук, генеральный директор ООО «Инфодор-Инфо» должность, ученая степень	Э.А. Цокур инициалы, фамилия
Нормоконтролер	 подпись, дата	зав. каф. ВПВ канд.техн.наук., доцент должность, ученая степень	Д.А. Кузьмин инициалы, фамилия

Красноярск 2020