



DriveHack

OFFLINE - ХАКАТОН

Распознавание сканов документов

Команда “В автобусе”

Участники команды



Андреичев Степан
Frontend



Иванов Владислав
Fullstack



Жданов Степан
Backend

Задачи

1. Автоматизация процесса классификации типа документа по его скану
2. Минимизация ошибок распознавания.

АО "АЛЬФА-БАНК", Москва

Б/К: 04452553

Сч. № 30101810200000000593

Б/К: 4070281010400003927

ИНН 7704003001

ОСОО "ИНТЕЛЛЕКТУАЛЬНЫЕ СОЦИАЛЬНЫЕ СИСТЕМЫ"

Получатель

Счет на оплату № 5 от 28 февраля 2020 г.

Поставщик: ООО "ИНТЕЛЛЕКТУАЛЬНЫЕ СОЦИАЛЬНЫЕ СИСТЕМЫ", ИНН 7704003001, КПП (Уполномочитель) 770101001, 121006, Москва г, Секретная Инновационного Центра тор, пр.Бела ул, дом № 7, ЭП/Пом/Иб 114603, тел.: +7 (495) 368-02-77

Получатель (Земельщик): Государственное унитарное предприятие города Москвы "Мосгортранс", ИНН 7706002402, КПП 770601001, 119035, Москва г, Реутовская наб, дом № 22/1, строение 1

Основание: Гражданско-правовой договор №69-СИ/19 от 19.12.2019г.

№	Товары (работы, услуги)	Кол-во	Ед.	Цена	Сумма
1	Оплата услуг за 2 квартал 2019г. Оплата услуг за 2 квартал 2019г. Оплата услуг за 2 квартал 2019г.	1	шт	2 714 427,41	2 714 427,41

Итого: 2 714 427,41

Без налога (НДС) 2 714 427,41

Всего к оплате: 2 714 427,41

Всего наименований 1, на сумму 2 714 427,41 руб.

Для исполнения суммы четырнадцать тысяч четыреста двадцать семь рублей 41 копейка

Оплата должна быть произведена в течение 10 рабочих дней.

Уведомление об оплате обязательно, в противном случае не производится возврата товара на складе.

Товар принимается по факту прихода денег на р/с Получателя, соответствием, при наличии доверенности и паспорта.

Руководитель:   

Бухгалтер: 

Климова Т. Р.

Всего № 2563

14.03.2020г.

Счёт на оплату



Решение

Реализовано 2 алгоритма классификации:

1. Tesseract OCR
2. Pytorch Convolutional neural network (VGG16)

Счет на оплату № 14 от 08 февраля 2016 г.



Счет на оплату №14 от 08 февраля 2016г.

Архитектура и стек технологий

Backend:

- Python
- Flask
- PyTorch
- Tesseract



Frontend:

- C#
- WPF
- httpClient



Как можно доработать

1. Добавить другие алгоритмы для классификации документов.
2. Расширить количество опознаваемых документов
3. Поддержка сканов в других форматах

[illegible]