



Восстановление пунктуации в тексте

Команда: SuperBytes

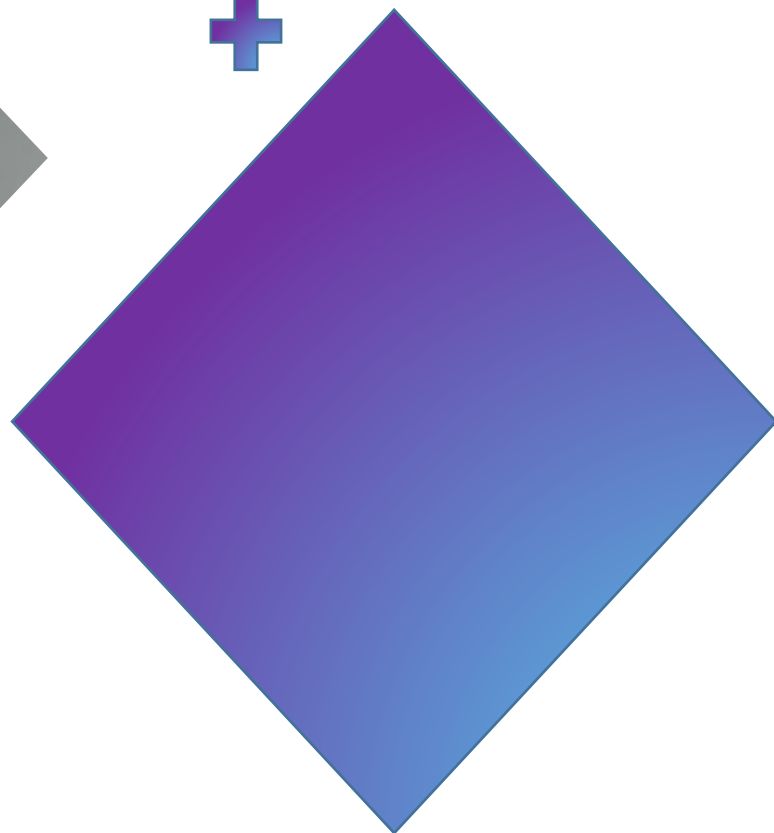
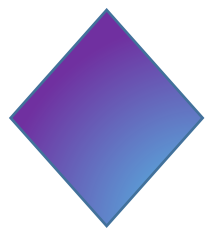
Космынина Валерия
Чернышёв Егор
Тарасов Георгий
Скребцов Никита
Шкварлюк Вадим



User story

Как пользователь, я хочу иметь возможность вставить текст в инструмент и получить текст с правильно расставленными знаками препинания, чтобы мне не приходилось тратить время на ручную проверку и исправление каждой ошибки.

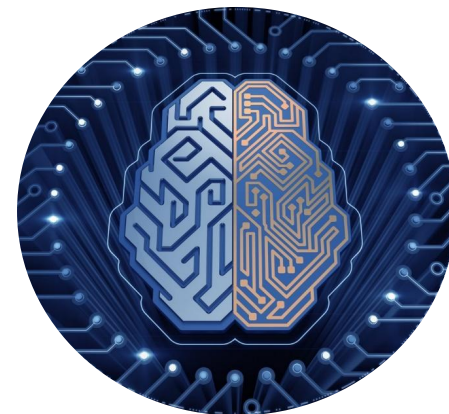
Проблематика



- Неправильно расставленные знаки препинания могут повлиять на восприятие и понимание письменного текста.
- При написании большого текста легко упустить какой-то знак препинания.

Решение

- 1) Данное решение предназначено для корректировки текста, удаляя неправильно расставленные знаки препинания из введенного текста и затем восстанавливая их на основе предварительно обученной модели. Рассмотрим сценарий обычного ввода текста:
- Пользователь вводит текст в текстовое поле.
- Приложение обрабатывает введенный текст, удаляя неправильно расставленные знаки препинания.
- Затем применяется обученная модель для восстановления знаков препинания в тексте.
- Восстановленный текст, с правильно расставленными знаками препинания, выводится на экран в отдельном поле.




- 2) Сценарий использования компьютерного зрения:
 - Пользователь нажимает на кнопку “Распознать текст с изображения”
 - Приложение обрабатывает изображение, выделяя текст и вставляет его в форму.
 - Затем применяется обученная модель для восстановления знаков препинания в тексте.
 - Восстановленный текст, с правильно расставленными знаками препинания, выводится на экран в отдельном поле.



Таким образом, данное решение помогает автоматически добавлять знаки препинания в текст, где они отсутствуют, облегчая чтение и понимание текста для пользователя.

Преимущества

- Наше решение не только определяет места, где должны находиться знаки препинания, но к тому же расставляют их.
- Наше решение учитывает контекст при расстановке знаков препинания.
- Наше решение применяет компьютерное зрение для распознавания текста с фотографии.

 Корректировка текста—□×

Введите текст без знаков пунктуации:

израильский министр по делам диаспоры нахман шай сообщил что пристально следит за развитием событий в американском городе колливиль штат техас где неизвестный захватил заложников в синагоге

Распознать текст с изображения

Обработать

Израильский министр по делам диаспоры нахман шай сообщил, что пристально следит за развитием событий в американском городе колливиль, штат техас, где неизвестный захватил заложников в синагоге



Стек технологий



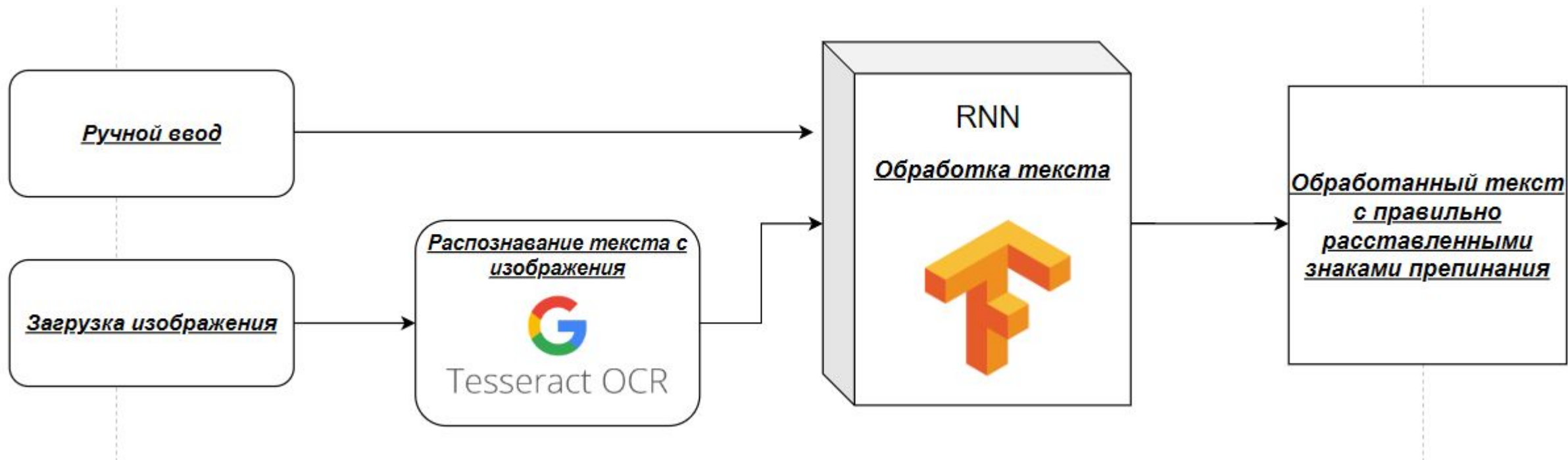
Tesseract OCR



NumPy



Архитектура



Метрики качества:

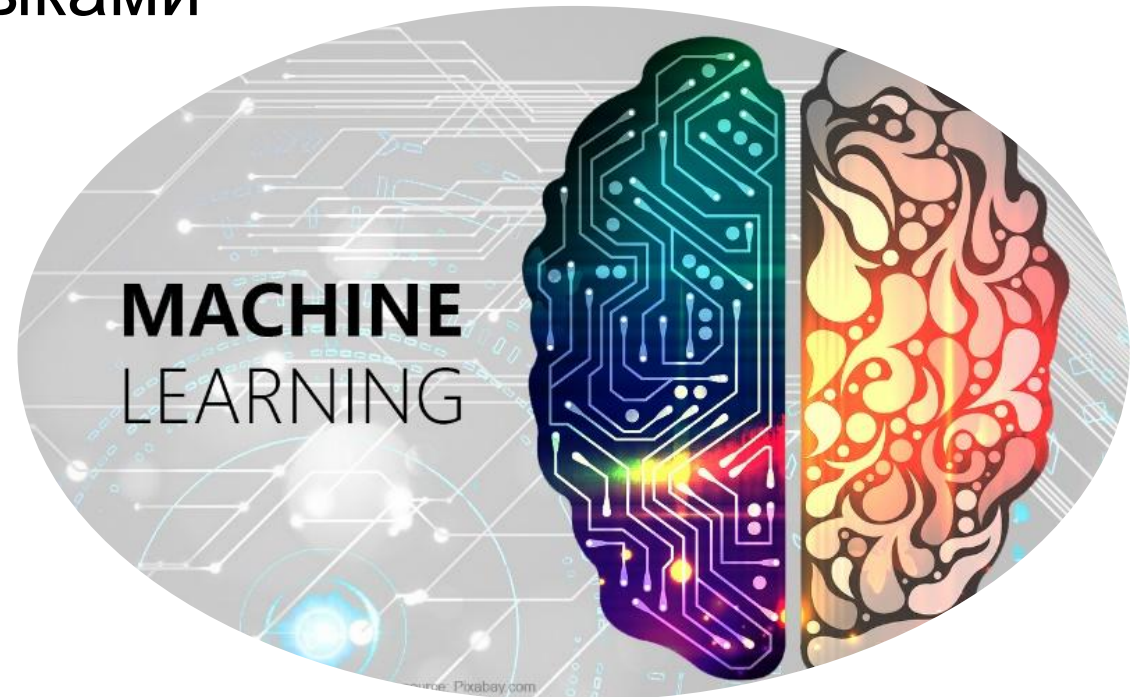
Знак	PRECISION	RECALL	F-SCORE
.	93.8	93.1	93.4
?	92.8	88.9	90.9
,	94.1	89.7	91.9

- Precision можно интерпретировать как долю объектов, названных классификатором положительными и при этом действительно являющимися положительными
- Recall показывает, какую долю объектов положительного класса из всех объектов положительного класса нашла нейросеть.
- F-SCORE — среднее гармоническое precision и recall



Планы на будущее

- Увеличить список знаков препинания
- Добавить работу с разными языками



SuperBytes



Скребцов Никита

<https://vk.com/skrippa>

Data scientist



Шкварлюк Вадим

<https://vk.com/id72206640>

Developer



Космылина Валерия

Team – lead
<https://vk.com/id462543175>



Тарасов Георгий

<https://vk.com/tearnasss>

Analytic



Чернышёв Егор

https://vk.com/egor_chernishev

Product manager



Ссылки:

GitHub :

<https://github.com/alloe16/punctuation-restoration/tree/main>

Команда:

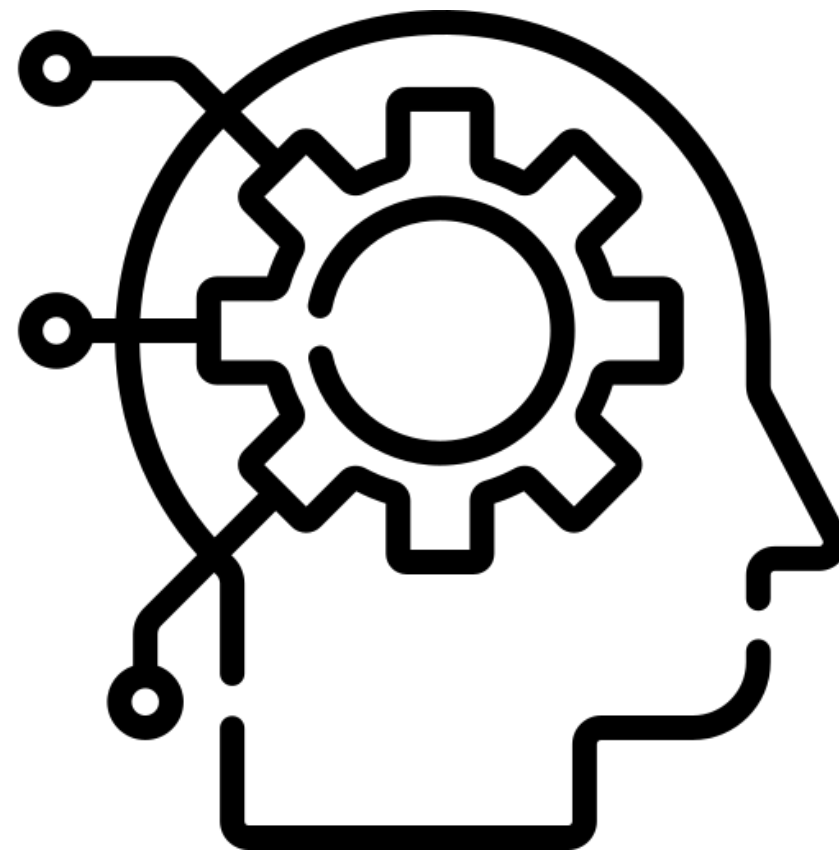
<https://vk.com/skrripa>

<https://vk.com/id462543175>

<https://vk.com/tearnasss>

<https://vk.com/id72206640>

https://vk.com/egor_chernishev



Спасибо за внимание!