**Московский государственный технический**

**университет им. Н.Э. Баумана**

Факультет «Информатика и системы управления»

Кафедра ИУ5 «Системы обработки информации и управления»

Курс «Парадигмы и конструкции языков программирования»

Отчет по лабораторной работе №3

«Обработка изображений и подготовка текста. Определение типов»

|  |  |  |
| --- | --- | --- |
| Выполнил: |  | Проверил: |
| студент группы ИУ5-32Б |  | преподаватель каф. ИУ5 |
| Яковлев Сергей |  | Гапанюк Ю.Е. |
| Подпись и дата: |  | Подпись и дата: |

Москва, 2024 г.

### Задание:

Разработать алгоритм обработки поступающих изображений и подготовки текста к анализу:

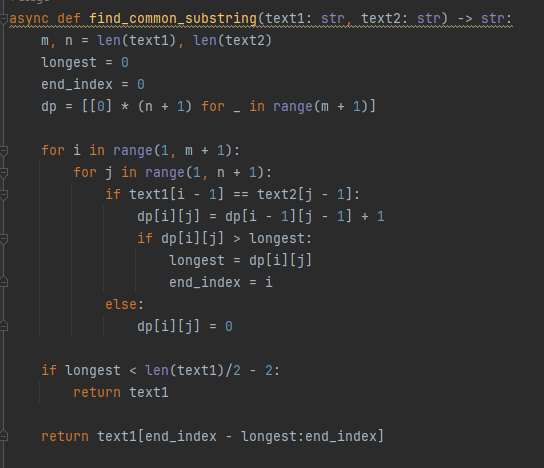
**Выполнение:**

Для обработки изображений было принято решение использовать формат PillowImage, а ля считывания EasyOCR. В предыдущем отчете изображение от пользователя было успешно получено. Однако, т.к. нельзя исключать неточности в считывании текста, а также особенности некоторых клейм (например: ЛОЗы зачастую дублируются в перевернутом формате) нужно произвести некоторые манипуляции с изображением.

1. Проверяем на содержание перевернутых элементов и отсеиваем таковые при наличии:

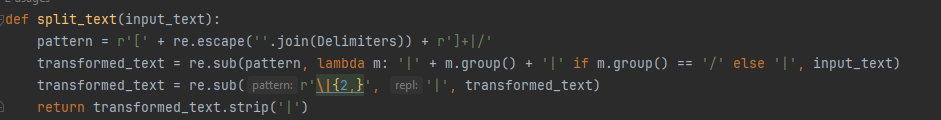


Выделяем наиболее длинную общую подстроку у двух вариантов текста, она и будет искомой:



Разделяем текст на части и преобразуем в массив сокращений:





Таким же образом разделим текст, поступающий сообщением.

Пример работы алгоритма с изображением, содержащим перевернутый текст:



Пример работы с текстом:



К сожалению, каждый тип клейм ввиду своих особенностей будет требовать собственных функций для его определения. Попытки использования регулярных выражений для упрощения процесса анализа ни к чему не привели.  
  