

Сервис распознавания содержимого платежных счетов

Полное название проекта

Сервис распознавания содержимого платежных счетов.

Цель

Реализовать сервис, предоставляющий функционал распознавания содержимого pdf-скана платёжного счёта. Интерфейсом взаимодействия с сервисом является ТГ-бот.

Результат (продукт)

Разработано приложение, содержащее:

- Серверную систему, предоставляющую HTTP-endpoint для загрузки pdf-файла со сканом платежного счета, возвращающий json-объект, который отображает структуру документа и хранит в соответствующих полях их содержимое. В случае неодноуровневой вложенности данных структура сохраняется в дочерних json-объектах. В случае невозможности распознавания значения какого-либо поля в результирующий json в соответствующее значение записывается установленное командой спецслово. В случае невозможности чтения наименования поля платежного счета это поле с его значением в результирующий json не включаются. Хранение истории запросов не предусматривается.
- Телеграм-бота, являющегося интерфейсом взаимодействия пользователя и серверной системы. Бот предоставляет единственную функцию загрузки pdf-файла, которая передает полученный файл серверной системе. В случае получения ответа со статусом 2xx/3xx json дублируется в текст ответного сообщения со специальным форматированием. В случае получения ответа 5xx в текст ответного сообщения записывается установленная командой спецфраза. В случае ответа 4xx в текст ответного сообщения записывается фраза из списка установленных командой на случай конкретной пользовательской ошибки.
- Jupyter-notebook(при необходимости несколько), демонстрирующий работоспособность выбранной модели и работоспособность общего пайплайна обработки файла с интерпретацией результатов работы модели.
- В случае обучения или дообучения модели Jupyter-notebook с обучением/дообучением.

Критерии приемки результата / продукта

- Приложение контейнеризовано и запускается консольной командой в изолированной среде.
- Пайплайн обработки pdf-документа корректно работает и это продемонстрировано в приложенном Jupyter-notebook'е.

- ТГ-бот и серверная часть корректно взаимодействуют без непредвиденных сбоев.
- ТГ-бот и серверная часть корректно реализуют все сценарии работы, представленные в описании продукта