

DAISY 2.0



Цели.

- Выделение договоров из набора файлов. Определение типа договора, контрагента, даты договора и периодов по договору, вида страхования и типа покрытия.
- Проверка корректности в сравнении с матрицей оговорок и другими версиями.
- Демонстрация найденных и отсутствующих в документе оговорок.
- Формирование отчётов.
- Формирование договоров.
- Исправление, частичное исправление, ручная корректировка.
- Дообучение моделей.

Фичи.

- Рисование областей распознавания вручную.
- Все системные шрифты + обучать новым шрифтам.
- Изменение форматирования текста.
- Варианты автозамены для плохо-распознанных символов.
- Документы, защищённые паролем.
- Создание и настройка сущностей.
- Сравнение версий документов как в гите.

Прототип.

- Несколько сканов документов.
- Простая матрица оговорок.
- Модуль распознавания документов (pytesseract).
- Модуль выявления оговорок и рекомендаций (определяем основные атрибуты, классифицируем, выявляем оговорки).
- Модуль разметки текстовых документов (просто выделяем простые сущности + статистика).
- Модуль верификации документов (опечатки).
- Примитивный UI (Jupyter notebook, R shiny web app).

Идём сверху-вниз смотрим, сколько успеем сделать к 15-му мая.