

ПОСТАНОВКА ЗАДАЧИ

НЕОБХОДИМО ДОБАВИТЬ ТОС В PDF-ФАЙЛЫ

ПРИМЕРЫ — ЭТО ДОКУМЕНТЫ, У КОТОРЫХ У КОТОРЫХ ЕСТЬ И ОГЛАВЛЕНИЕ И БОКОВАЯ ПАНЕЛЬ.

ТРЕБОВАНИЯ К РЕШЕНИЮ:

1. НАПИСАТЬ ПРОГРАММУ, ДОБАВЛЯЮЩУЮ ТОС В PDF ФАЙЛ;

РЕЗУЛЬТАТ: GIT-РЕПОЗИТОРИЙ

2. ПОДГОТОВИТЬ ДОКЛАД С ОПИСАНИЕМ МЕТОДИКИ ДОБАВЛЕНИЯ ТОС.

РЕЗУЛЬТАТ: ПРЕЗЕНТАЦИЯ

ДОПОЛНИТЕЛЬНАЯ ЗАДАЧА:

РАЗРАБОТАТЬ ПОДХОД К ФОРМИРОВАНИЮ ТОС В ДОКУМЕНТАХ, ГДЕ РАЗДЕЛ СОДЕРЖАНИЕ ИЗНАЧАЛЬНО ОТСУТСТВУЕТ

ПЛАН РАЗРАБОТКИ

ПЛАН:

1. ПОДКЛЮЧИТЬ OCR И ОБРАБОТАТЬ ТЕКСТ С ОГЛАВЛЕНИЯ;
2. ЕСЛИ ОГЛАВЛЕНИЕ ОТСТУТСТВУЕТ ТО ВОСПОЛЬЗОВАТЬСЯ LLM ЧТОБЫ ПОЛУЧИТЬ ОГЛАВЛЕНИЕ;
3. ПОДНЯТЬ АПИ;
4. ПОДНЯТЬ ФРОНТ;
5. АРЕНДОВАТЬ СЕРВЕР И ПОДНЯТЬ ФИНАЛЬНУЮ ВЕРСИЮ НА НЕМ.

ТАКЖЕ КАЖДЫЙ ЭТАП КОММИТИТЬ В GIT-РЕПОЗИТОРИЙ

РЕАЛИЗАЦИЯ

ДЕНЬ 1. ВЕЧЕР ВТОРНИКА

РЕСЕРЧ OCR РЕШЕНИЙ: [PADDLEOCR](#) ПОКАЗАЛ НЕ САМУЮ ЛУЧШУЮ ТРАНСКРИБАЦИЮ, [TESSERACTOCR](#) ЖЕ ПОКАЗАЛ СЕБЯ НАМНОГО ЛУЧШЕ.

ИЗ ПРЕДОБРАБОТКИ ИСПОЛЬЗОВАЛ [GRAYSCALE](#), [BILATERAL FILTER](#), И БИНАРИЗАЦИЮ С ПОМОЩЬЮ [МЕТОДА ОЦУ](#).
ЗАМЕТИВ ПОСТОЯННЫЕ ОШИБКИ OCR, ЗАМЕНИЛ СИМВОЛЫ [\\$](#), [&](#) НА [8](#) И ТД. НА ПОСТОБРАБОТКЕ .
С ПОМОЩЬЮ РЕГУЛЯРНЫХ ВЫРАЖЕНИЙ ОПРЕДЕЛЯЛ НОМЕР СТРАНИЦЫ.

ДЕНЬ 2. ВЕЧЕР СРЕДЫ

[ПОДНИМАЯ LLAMA](#), Я СТОЛКНУЛСЯ С ПРОБЛЕМОЙ НЕДОСТУПНОСТИ МОДЕЛИ В СТРАНЕ, ПОЭТОМУ БЫЛА ИСПОЛЬЗОВАНА [YANDEXGPT API](#). ПОПОЛНИВ СЧЕТ, Я ПОДКЛЮЧИЛ [YANDEXGPT API](#) И ЗАНЯЛСЯ ПРОМПТ-ИНЖИНЕРИНГОМ.

РЕАЛИЗАЦИЯ

ДЕНЬ 3. ВЕЧЕР ЧЕТВЕРГА

К ЭТОМУ ВРЕМЕНИ БЫЛ РЕАЛИЗОВАН ФРОНТ.

Я НАЧАЛ ПИСАТЬ АПИ НА [FASTAPI](#), НО СТОЛКНУЛСЯ С ПРОБЛЕМОЙ ОБРАБОТКИ PDF-ФАЙЛОВ И СОЗДАНИЕМ КАНВАСА ДЛЯ ТОГО, ЧТОБЫ ВЫПОЛНИТЬ ДОПОЛНИТЕЛЬНОЕ ЗАДАНИЕ.

ДАЛЕЕ Я СТОЛКНУЛСЯ С ПРОБЛЕМОЙ ДОЛГОЙ ОБРАБОТКИ ФАЙЛА, ИЗ-ЗА ЧЕГО ФРОНТ ВЫДАВАЛ ОШИБКУ.

ПЕРЕХОД НА [GRADIO](#) РЕШИЛ ЭТУ ПРОБЛЕМУ. РАЗВЕРНУВ ЕГО, НА ЭТОМ ЗАКОНЧИЛ.

К СОЖАЛЕНИЮ, ОТ ИЗНАЧАЛЬНОГО ФРОНТА ПРИШЛОСЬ ОТКАЗАТЬСЯ.

ДЕНЬ 4. ВЕЧЕР ВОСКРЕСЕНЬЯ

ПРОВЕРИЛ КОРРЕКТНОСТЬ РАБОТЫ СЕРВИСА И ПОФИКСИЛ НЕКОТОРЫЕ БАГИ, СДЕЛАЛ ПРЕЗЕНТАЦИЮ С ОПИСАНИЕМ ПРОДЕЛАННОЙ РАБОТЫ.

РЕЗУЛЬТАТ И UI

ЧТО МОЖЕТ СЕРВИС?

СЕРВИС МОЖЕТ СОЗДАВАТЬ БОКОВУЮ ПАНЕЛЬ для PDF-ФАЙЛОВ, У КОТОРЫХ ЕСТЬ ОГЛАВЛЕНИЕ.

ТАКЖЕ ОН МОЖЕТ СОЗДАВАТЬ ОГЛАВЛЕНИЕ для ФАЙЛОВ, У КОТОРЫХ ЕГО НЕТ.

НЕДОСТАТКИ

ОГЛАВЛЕНИЕ НЕ ВСЕГДА ПОЛУЧАЕТСЯ КОРРЕКТНОГО ФОРМАТА.

ПРИ УКАЗАНИЕМ НЕКОРРЕКТНОГО НОМЕРА, СИСТЕМА НЕ СМОЖЕТ ОБРАБОТАТЬ ЭТО КАК ОШИБКУ

ПРЕДЛАГАЕМОЕ РЕШЕНИЕ

- ИСПОЛЬЗОВАНИЕ **CHATGPT** ИЛИ **LLAMA**;
- ВЫДЕЛЕНИЕ БОЛЬШЕГО ВРЕМЕНИ НА **ПРОМПТ-ИНЖИНЕРИНГ**;
- ВНЕДРЕНИЕ **СИСТЕМЫ ПОИСКА** ОГЛАВЛЕНИЯ В ФАЙЛЕ.

РЕЗУЛЬТАТ И UI

КАК РАБОТАЕТ?

1. ЗАГРУЖАЕМ ФАЙЛ И НАЖИМАЕМ КНОПКУ "ОБРАБОТАТЬ"

- **DOS_TYPE** ФАЙЛЫ КОТОРЫЕ ИМЕЮТ БОКОВУЮ ПАНЕЛЬ И ОГЛАВЛЕНИЕ, НЕ ВАЖНО КАКУЮ СТРАНИЦУ УКАЖЕТ ПОЛЬЗОВАТЕЛЬ, СИСТЕМА ВЕРНЕТ ЕГО ЖЕ ФАЙЛ, ТАК КАК ОН УЖЕ НУЖНОГО ФОРМАТА
- **PAGE_TYPE** ФАЙЛЫ КОТОРЫЕ ИМЕЮТ ОГЛАВЛЕНИЕ НО НЕ ИМЕЮТ БОКОВУЮ ПАНЕЛЬ, ПОЛЬЗОВАТЕЛЬ ДОЛЖЕН УКАЗАТЬ КОРРЕКНУТЮ СТРАНИЦУ С ОГЛАВЛЕНИЕМ
- **ОТСУТСТВУЕТ** - ФАЙЛ НЕ ИМЕЕТ ОГЛАВЛЕНИЯ И БОКОВОЙ ПАНЕЛИ, ПОЛЬЗОВАТЕЛЮ НЕ НУЖНО УКАЗЫВАТЬ НИКАКУЮ СТРАНИЦУ

2. ПОСЛЕ ОБРАБОТКИ, ПОЛЬЗОВАТЕЛЬ УВИДИТ ГОТОВЫЙ ФАЙЛ, И ЕСЛИ ОН ЕГО УСТРОИТ, ТО НИЖЕ ОН СМОЖЕТ СКАЧАТЬ ФАЙЛ

Upload a PDF

Drop PDF

Click to Upload

Выберите действие

doc_type

page_type

Отсутствует

Страница с оглавлением, если присутствует. Иначе любое число

0

Output

Drop PDF

Click to Upload

Скачать файл

Обработать

Оглавление

Раздел 1

Статья 1. Общие положения 3

Статья 2. Фирменное наименование 3

Статья 3. Ответственность 3

Статья 4. Филиалы и представительства 4

Статья 5. Структура Общества. Участие Общества в коммерческих организациях 4

Раздел 2

Статья 6. Цели и предмет деятельности Общества 4

Раздел 3

Статья 7. Уставный капитал 7

Статья 8. Размещение Общественных акций. Приобретение Общественных размещенных акций 8

Статья 9. Фонды и чистые активы Общества 8

Раздел 4

Статья 10. Права и обязанности акционеров 9

Статья 11. Распределение прибыли. Дивиденды Общества 11

Статья 12. Реестр акционеров 11

Раздел 5

Статья 13. Органы управления Общества. Ответственность членов Совета директоров, генерального директора Общества 12

Статья 14. Общее собрание акционеров. Общие положения 12

Статья 15. Компетенция Общего собрания акционеров 13

Статья 16. Порядок принятия решений Общим собранием акционеров. Доведение решений Общего собрания акционеров до сведения акционеров 14

Статья 17. Информация о проведении Общего собрания акционеров 15

Статья 18. Повестка дня Общего собрания акционеров 15

Статья 19. Подготовка к проведению Общего собрания акционеров 17

Статья 20. Внеочередное Общее собрание акционеров 18

Статья 21. Иные вопросы, связанные с проведением Общего собрания акционеров 19

Статья 22. Совет директоров Общества 20

Статья 23. Компетенция Совета директоров Общества 20

Статья 24. Избрание Совета директоров Общества 24

Статья 25. Структура Совета директоров Общества 24

Статья 26. Заседание Совета директоров Общества 24

Статья 27. Единственный исполнительный орган Общества - Генеральный директор 25

Статья 28. Компетенция Генерального директора Общества 26

Статья 29. Ответственность членов Совета директоров Общества, Генерального директора 28

Раздел 6

Статья 30. Аудитор Общества 28

Статья 31. Учет и отчетность. Документы Общества 28

Статья 32. Реорганизация и ликвидация Общества 29

2

ЗАТРАТЫ И ВЫВОД

ЗАТРАТЫ СОСТАВИЛИ:

1. 2700 РУБЛЕЙ ЗА АРЕНДУ СЕРВЕРА НА МЕСЯЦ;
2. 300 РУБЛЕЙ ЗА ПРОМПТЫ ВО ВРЕМЯ ТЕСТИРОВАНИЯ И РАЗРАБОТКИ СИСТЕМЫ;
3. ЗАТРАТЫ НА ЗАРПЛАТУ ДЛЯ СОТРУДНИКА.

ОБЩИЕ ЗАТРАТЫ:

3000 РУБЛЕЙ + ЗП СОТРУДНИКА ЗА 2 ДНЯ РАБОТЫ