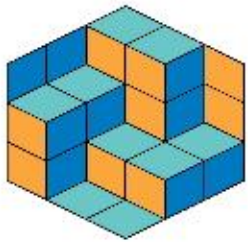


Бот по распознаванию протоколов

СПБГУ, МКН

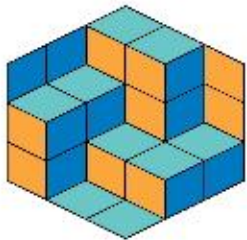
Александр Комнацкий, Александр Наумцев, Кузнецов Кирилл

Руководитель: Авдюшенко Александр Юрьевич



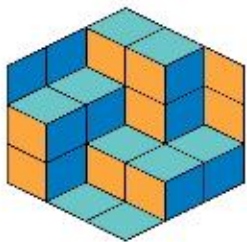
О проекте

- Существует множество научных конкурсов с оценочной системой.
- Результатов очень много.
- Приходится вручную переносить большие данные.



Задачи

- Необходимо переносить оценки с печатных бланков в электронный формат
- Научиться распознавать
- Написать бота
- Выбрать способ сохранять информацию

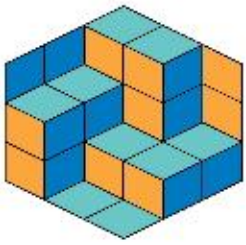


Пример бланка

Протокол оценки конкурсных работ по секции «Химия» научного жюри.
XVIII Балтийский научно-инженерный конкурс, 24–26 марта 2022 г.

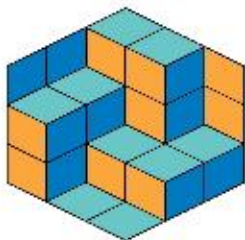
СН

№	Название работы	Группа (1-5)	Соответствие целей результатам	Новизна, оригинальность	Актуальность, востребованность	Сложность	Методы и технологии	Трудоемкость и ресурсоёмкость	Законо- ченность	Комму- никативная культура	Комментарий
СН01	Сравнение методов пробоподготовки при определении подвижных форм металлов (2 автора)	2	4	8	9	1	3	7	6	5	
СН02	Новые кристаллические формы фармацевтической соли ципрофлоксацин - 4-аминобензойная кислота (1 автор)	1	7	9	2	6	5	9	1	2	
СН03	Приготовление безалкогольного пива с применением безглютенового солода (2 автора)	5	6	7	8	9	3	2	4	7	
СН04	Водный синтез нанопорошков ZnO в присутствии SDS и n-PnOH для фотокаталитического разложения метиленового синего (1 автор)	1	3	8	7	9	6	8	5	9	
СН05	Разработка и применение нейросети для анализа изображений реакционно-диффузионных систем (1 автор)	1	3	2	8	0	7	4	8	1	



Распознавание таблиц

- Распознавание таблиц
 - Применяем пороговую обработку
 - OpenCV (морфологические операции)
- Найти блоки клеток
- Распознать содержимое клеток

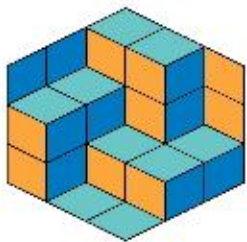


После операций

Протокол оценки конкурсных работ по секции «Химия» научного жюри.
XVIII Балтийский научно-инженерный конкурс, 24–26 марта 2022 г.

СН

№	Название работы	Группа (1-5)	Соответствие целей результатам	Новизна, оригинальность	Актуальность, востребованность	Сложность	Методы и технологии	Трудоемкость и ресурсоемкость	Заключенность	Коммуникативная культура	Комментарий
СН01	Сравнение методов пробоподготовки при определении подвижных форм металлов (2 автора)	2	4	8	9	1	3	7	6	5	
СН02	Новые кристаллические формы фармацевтической соли ципрофлоксацин - 4-аминобензойная кислота (1 автор)	1	7	9	2	6	5	9	1	2	
СН03	Приготовление безалкогольного пива с применением безглютенового солода (2 автора)	5	6	7	8	9	3	2	4	7	
СН04	Водный синтез нанопорошков ZnO в присутствии SDS и n-PnOH для фотокаталитического разложения метиленового синего (1 автор)	1	3	8	7	9	6	8	5	9	
СН05	Разработка и применение нейросети для анализа изображений реакционно-диффузионных систем (1 автор)	1	3	2	8	0	7	4	8	1	

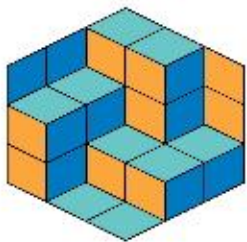


Выделение нужных прямых

Протокол оценки конкурсных работ по секции «Химия» научного жюри.
XVIII Балтийский научно-инженерный конкурс, 24–26 марта 2022 г.

СН

№	Название работы	Группа (1-5)	Соответствие целей результатам	Новизна, оригинальность	Актуальность, востребованность	Сложность	Методы и технологии	Трудоемкость и ресурсоемкость	Завершенность	Коммуникативная культура	Комментарий
СН01	Сравнение методов пробоподготовки при определении подвижных форм металлов (2 автора)	2	4	8	9	1	3	7	6	5	
СН02	Новые кристаллические формы фармацевтической соли ципрофлоксацин - 4-аминобензойная кислота (1 автор)	1	7	9	2	6	5	9	1	2	
СН03	Приготовление безалкогольного пива с применением безглютенового солода (2 автора)	5	6	7	8	9	3	2	4	7	
СН04	Водный синтез нанопорошков ZnO в присутствии SDS и n-PhOH для фотокаталитического разложения метиленового синего (1 автор)	1	3	8	7	9	6	8	5	9	
СН05	Разработка и применение нейросети для анализа изображений реакционно-диффузионных систем (1 автор)	1	3	2	8	0	7	4	8	1	

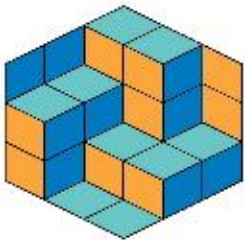


Пересечения

Протокол оценки конкурсных работ по секции «Химия» научного жюри.
XVIII Балтийский научно-инженерный конкурс, 24–26 марта 2022 г.

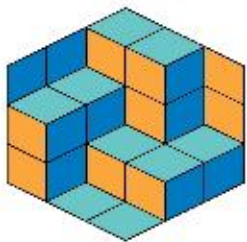
СН

№	Название работы	Группа (1-5)	Соответствие целей результатам	Новизна, оригинальность	Актуальность, востребованность	Сложность	Методы и технологии	Трудоемкость и ресурсоемкость	Заочность	Коммуникативная культура	Комментарий
СН01	Сравнение методов при определении подвижных форм металлов (2 автора)	2	4	8	9	1	3	7	6	5	
СН02	Новые кристаллические формы фармацевтической соли ципрофлоксацин - 4-аминобензойная кислота (1 автор)	1	7	9	2	6	5	9	1	2	
СН03	Приготовление безалкогольного пива с применением безглютенового солода (2 автора)	5	6	7	8	9	3	2	4	7	
СН04	Водный синтез нанопорошков ZnO в присутствии SDS и n-PnOH для фотокаталитического разложения метиленового синего (1 автор)	1	3	8	7	9	6	8	5	9	
СН05	Разработка и применение нейросети для анализа изображений реакционно-диффузионных систем (1 автор)	1	3	2	8	0	7	4	8	1	



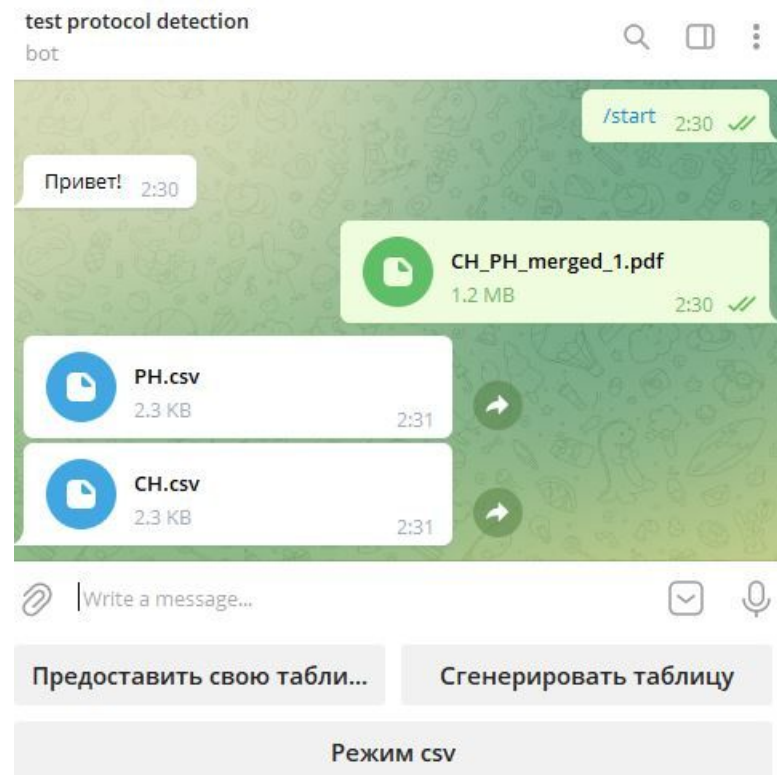
Распознавание ячеек

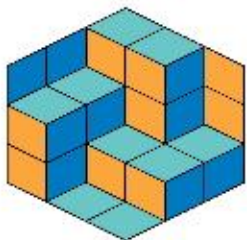
- Pytesseract для извлечения печатного текста
(плохо работает с рукописными цифрами)
- MNIST модель для извлечения рукописных чисел
(трудность: принимает особый формат чисел)



Бот

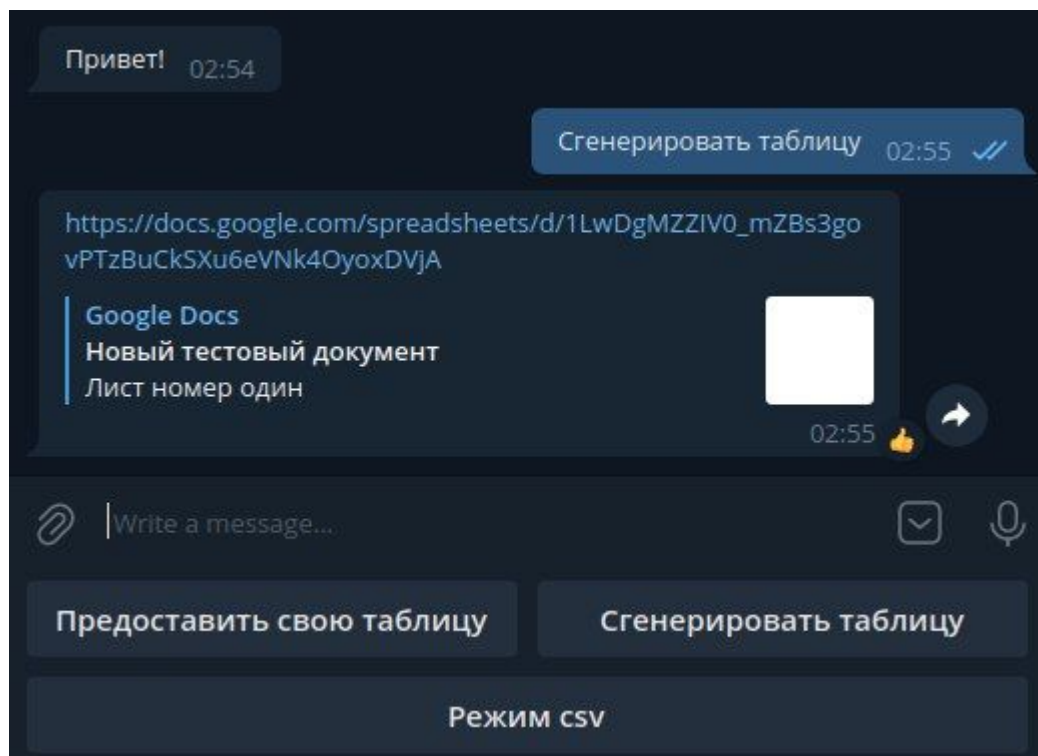
- Принимает на вход pdf с бланками
- Возвращает csv файлы по секциям
- Создает онлайн таблицу при подвязке гугла



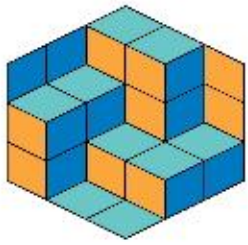


Google Sheets

- Написано удобное использование apiclient

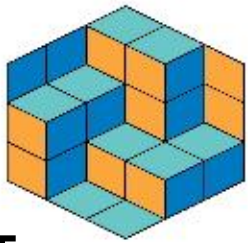


[illegible]



Итог

- Распознавание протоколов
- 1 лист обрабатывается 1.8 секунды
- Бот для работы с пользователем
- Удобный формат хранения
- https://github.com/bicikar/protocol_detection_bot



Технологии

- Pytesseract, OpenCV
- Telebot, json, csv
- apiclient (Sheets Api), http2lib, oauth2client