

Руководство пользователя к приложению по  
извлечению табличных данных из .pdf файлов,  
основанных на сканах

**Разработчики:**

Приходько Роман

Доронкина Анна

Королёв Никита

### **Описание решаемой задачи:**

1. Чтение файлов формата .pdf и отображение их содержимого на экране
2. Захват всех таблиц внутри этого файла
3. Реализация возможности экспорта таблиц в формате .xlsx или .csv
4. Поддержка подстройки параметров OCR приложения, если качество низкое

### **Технические требования:**

64-битная операционная система Windows, на которую возможна установка интерпретатора Python 3.8. Также приложение должно исполняться на компьютере средней мощности с 8 Gb оперативной памяти и 2Gb видеопамяти не очень мощной видеокарты.

### **Инструкция по установке:**

Пользователю требуется установить на свой компьютер дистрибутив Anaconda, и библиотеки, которые используются в программе.

### **Инструкция перед ПЕРВЫМ запуском программы:**

Пользователю необходимо открыть cmd – командную строку и установить необходимые библиотеки.

cd..

pip install camelot-py

pip install tabula-py

pip3 install pyqt5

pip install PyQt5-tools

Теперь пользователь может запустить программу, чтобы приступить к работе после запуска программы нажмите кнопку «Import» и выберите pdf файл. В противном случае кнопка «Import» не работает.

### **Для каждого запуска:**

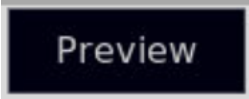

Так как необходимые библиотеки уже установлены, пользователь должен просто запустить программу и аналогично после нажатия кнопки «Import» выбрать pdf файл.

## Описание интерфейса:

При запуске программы открывается окно с кнопкой «Import».



После нажатия кнопки «Import» и выбора pdf файла пользователю предоставляется выбор:

-  Кнопка «Preview» показывает файл в формате pdf, который можно прочитать.
-  Кнопка «Export» предлагает пользователю выбрать удобную библиотеку для изъятия табличных данных из предоставленного pdf файла.

**После ее нажатия появятся две кнопки:**

1. 
2. 

Библиотека tabula-ру лучше подходит для извлечения простых таблиц, в то время как camelot-ру имеет большую вариативность настроек, которые позволяют лучше выделить «сложную» таблицу.