**Описание функциональных характеристик программы для ЭВМ**

**«Интеллектуальный модуль для классификации документов»**

Москва 2025 г.

Содержание

[1. Введение 3](#_Toc188368741)

[2. Назначение и условия применения 4](#_Toc188368742)

[2.1 Виды деятельности 4](#_Toc188368743)

[2.2 Программные и аппаратные требования к системе 4](#_Toc188368744)

[3. Состав системы 4](#_Toc188368745)

[3.1 Реализованные на данный момент базовые функции 4](#_Toc188368746)

[3.2 Полный функционал 5](#_Toc188368747)

[4. Функционал системы 5](#_Toc188368748)

[5. Эксплуатация системы 5](#_Toc188368749)

[5.1 Подготовка к работе 5](#_Toc188368750)

[5.2 Использование ИС по назначению 5](#_Toc188368751)

[5.3 Завершение работы ИС 10](#_Toc188368752)

[5.4 Аварийные ситуации 11](#_Toc188368753)

# **1. Введение**

Программа для ЭВМ «Интеллектуальный модуль для классификации документов» является средством для автоматизации классификации судебных документов с использованием машинного обучения.

Основные функции: классификация судебных документов на основе их содержания; определение типа документа; автоматическое перемещение документа в соответствующую папку в зависимости от класса. Программа предназначена для использования в коллекторской системе и может быть интегрирована с существующими системами документооборота. Программа обеспечивает высокую точность классификации судебных документов (достигая accuracy 0.976 на тестовой выборке), что позволяет ускорить и упростить процесс классификации и регистрации судебных документов. Программа также может быть использована в других областях, где требуется классификация документов на основе их содержания.

# **2. Назначение и условия применения**

## **2.1 Виды деятельности**

Программа относится к области вычислительной техники для выявления классификации судебных документов. Технический результат заключается в повышении точности классификации судебных документов.

Компьютерно-реализуемый способ содержит этапы, на которых:

- получают данные в текстовом формате преобразованного из представленных .pdf сканов;

- осуществляют обработку полученных данных с помощью алгоритмов машинного обучения, в ходе которой каждому тексту присваивается тег, соответствующий заданному типу документа;

- выполняют перемещение документа в соответствующую папку в зависимости от класса.

## **2.2 Программные и аппаратные требования к системе**

Языки программирования, применявшиеся при разработке ПО:

* Python 3.11

Среда разработки ПО:

* Jupyter Notebook.

Для корректной работы с платформой необходима следующая конфигурация автоматизированного рабочего места пользователя:

* Минимальные требования к системе - 4ядра
  + 4Gb RAMM доступной памяти
  + 100Gb SSD

Поддерживаемые ОС:

* + Windows 10

Необходимое ПО сторонних производителей:

* Tesseract-OCR
* Poppler-21.03.0

# **3. Состав системы**

## **3.1 Реализованные на данный момент базовые функции**

* Обработка PDF-документов.
* Распознавание текста в PDF-документах.
* Классификация документов по категориям.
* Сортировка документов по категориям.

## **3.2 Полный функционал**

* Обработка PDF-документов.
* Распознавание текста в PDF-документах.
* Классификация документов по категориям.
* Сортировка документов по категориям.
* Возможность добавления новых категорий и правил классификации.

# **4. Функционал системы**

Программа предназначена для автоматизации классификации судебных документов с использованием машинного обучения. Модуль работает под управлением ОС Windows и может быть легко адаптирован для работы на других операционных системах.

Программа состоит из следующих компонентов:

* + Модуль обработки PDF-файлов
  + Модуль обученной модели классификации текстов
  + Модуль сортировка документов по категориям

Среда разработки ПО:

* Изолированная подсеть на основе ОС Windows.

Для корректной работы с платформой необходима следующая конфигурация автоматизированного рабочего места пользователя:

* Минимальные требования к системе - 4ядра
  + 4Gb RAMM доступной памяти
  + 100Gb SSD

Действия по настройке и установке программного обеспечения не требуют специальных технических знаний и могут быть выполнены пользователем самостоятельно.

# **5. Эксплуатация системы**

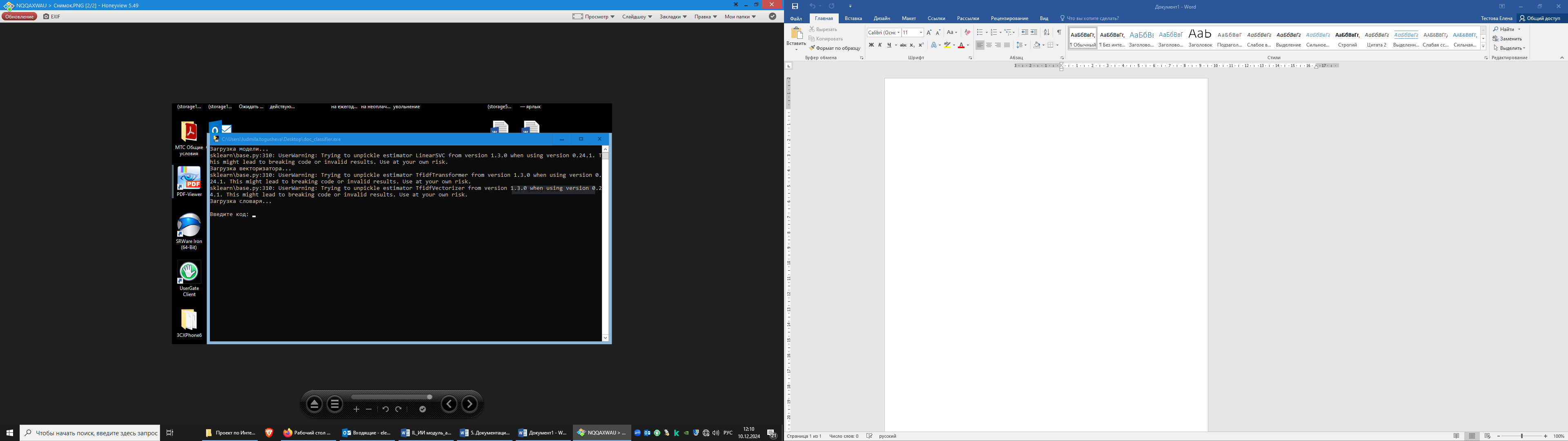
## **5.1 Подготовка к работе**

* Для начала работы требуется установить необходимое ПО сторонних производителей
* Для работы с программой необходимо иметь доступ к PDF-документам

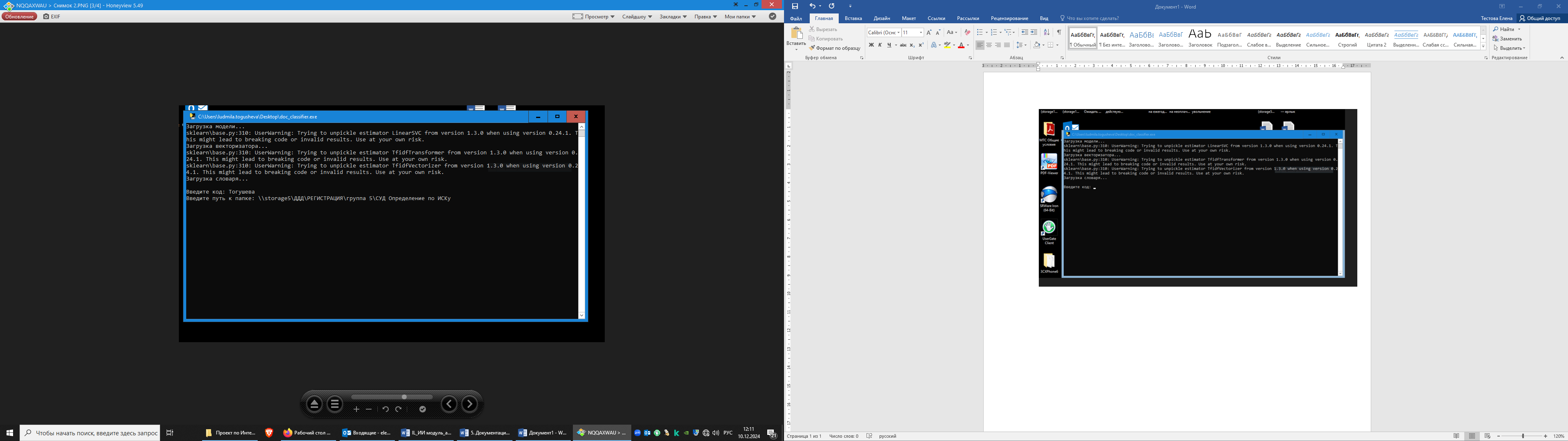
## **5.2 Использование ИС по назначению**

Метод представляет из себя следующее:

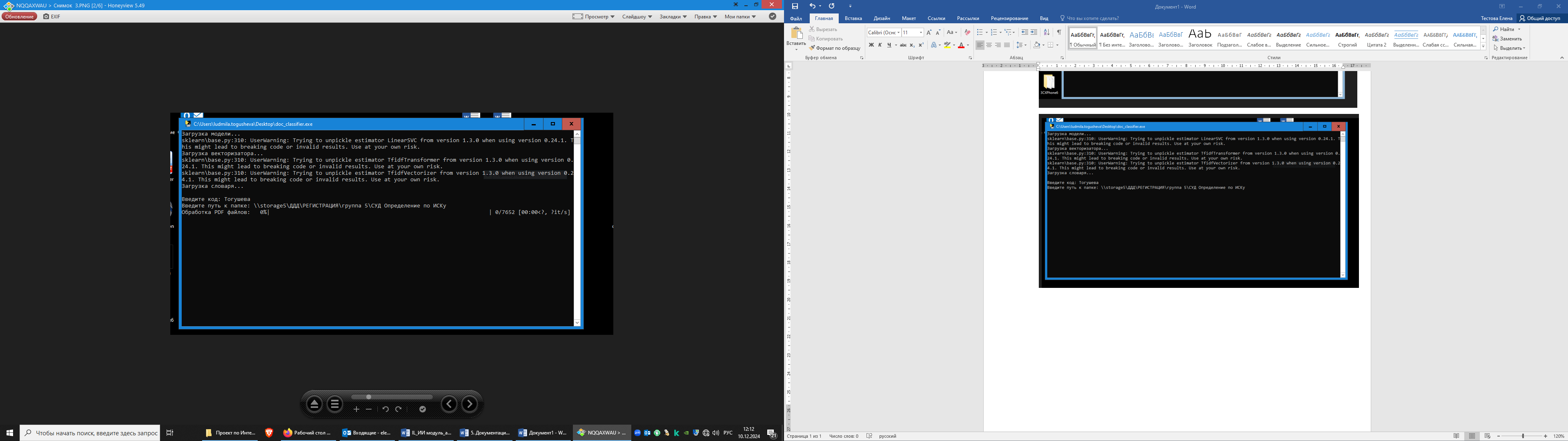
* Ввод кода сотрудника, обрабатывающего PDF документов

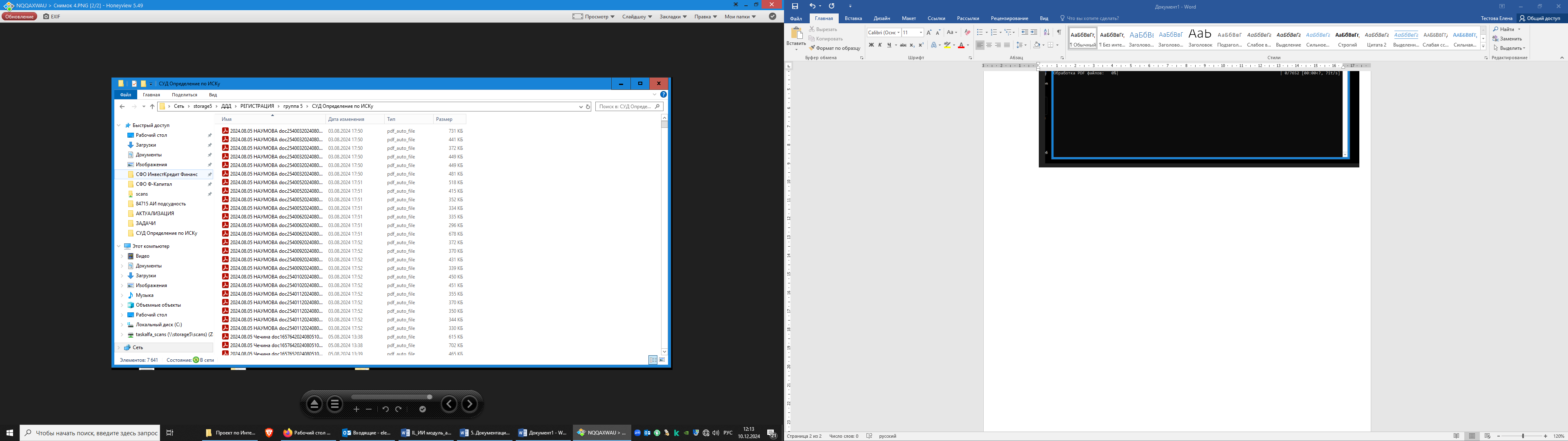


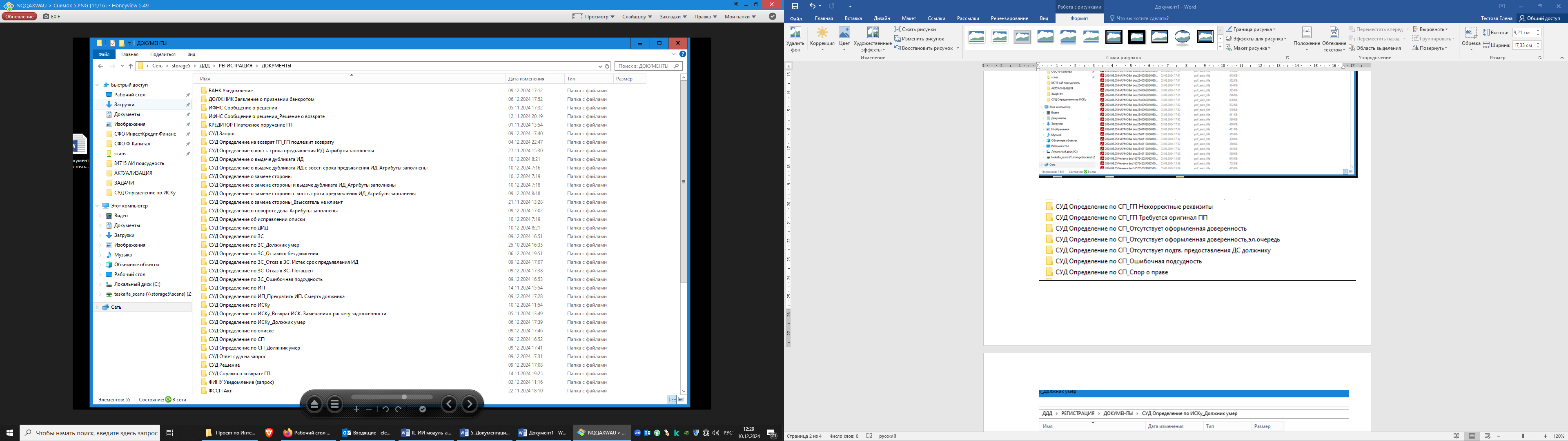
* Ввод ссылки папки, обрабатываемых PDF документов

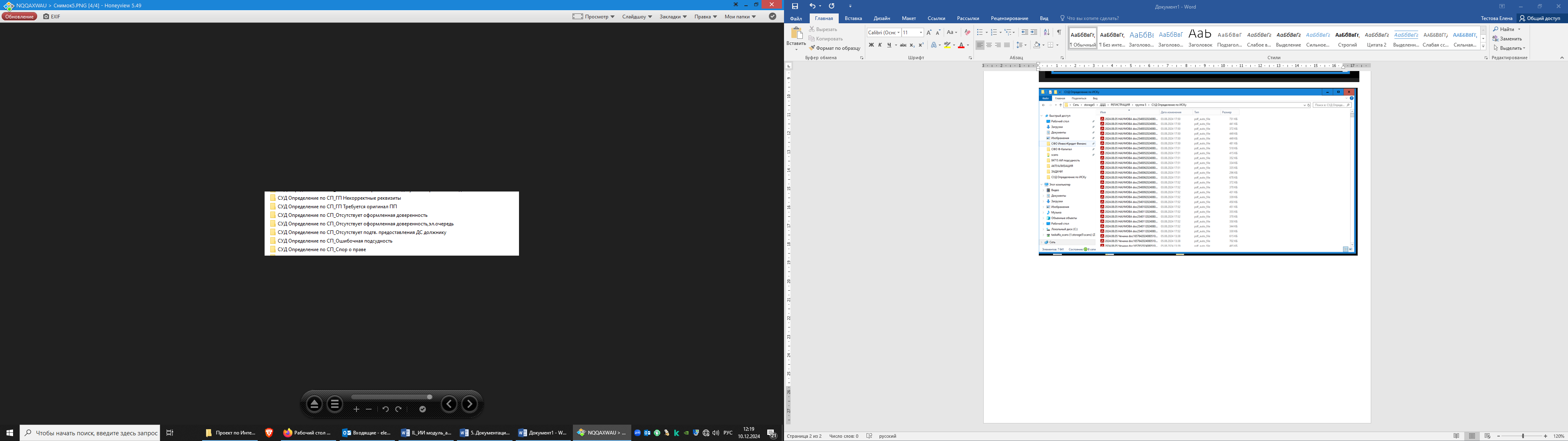


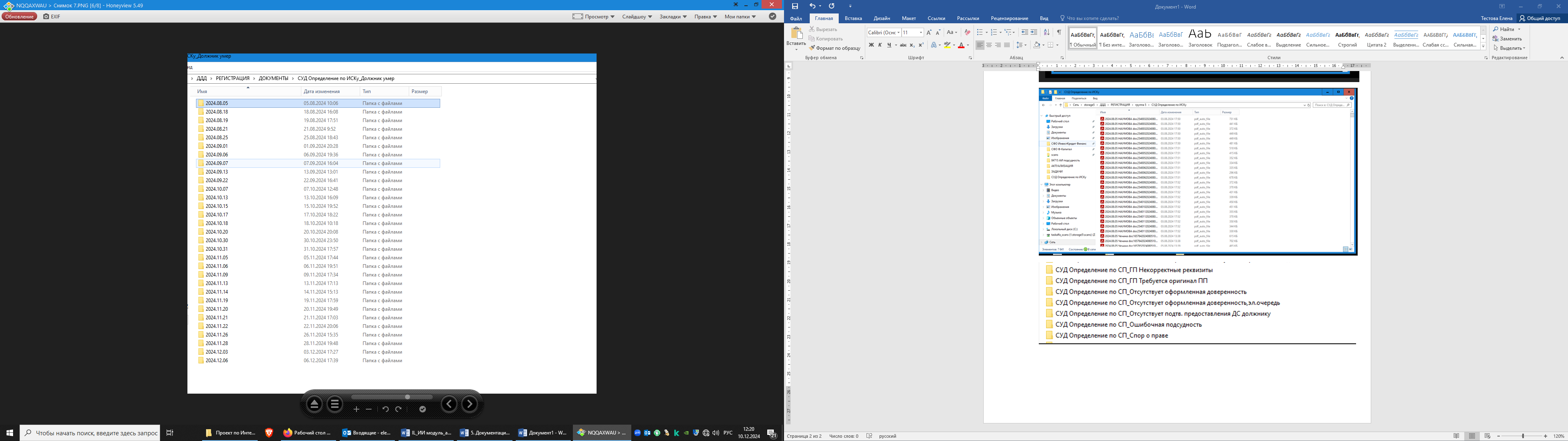
* Загрузка с последующей обработкой и сортировкой PDF-документов

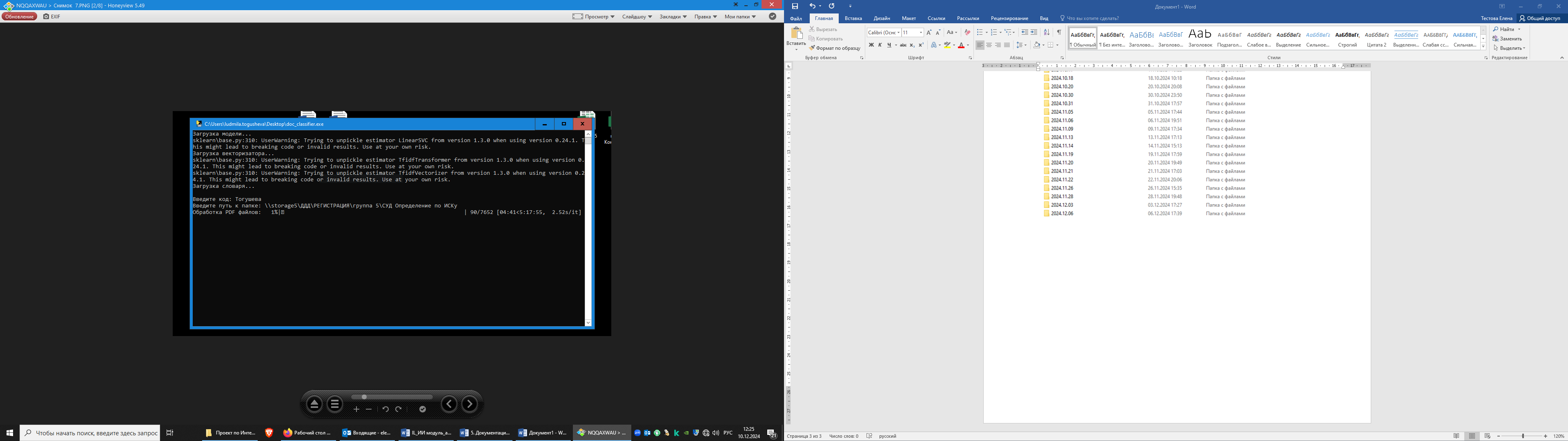






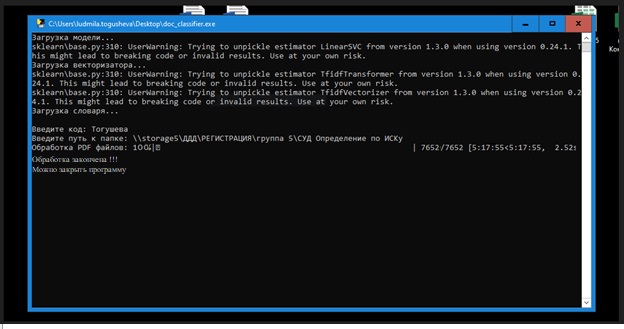






## **5.3 Завершение работы программы**

Программа завершает работу после обработки всех PDF документов:



## **5.4 Аварийные ситуации**

1. Информацию об аварийных ситуациях можно получить через лог-файлы программы
2. В случае ошибок в работе программы необходимо обратиться к разработчикам