

Руководство пользователя.

Программа SOICA Creator.

Версия 0.8.0

Утверждено:

Горюнов В.Н.

Генеральный директор ООО «Сойка»

Москва
28.07.2023 г.

Оглавление

1. Описание системы	4
2. SOICA Creator	5
2.1 Вход	5
2.2 Главный экран программы	5
3. Интерфейс проекта	14
3.1 Навигационное меню	15
3.2 Панель работы со страницей документа	26
3.3 272719	
3.4 270 свойствами объекта в документе	20
4. Профили распознавания	23
5. Профили экспорта	28
6. Системные требования программы	412
7. Установка программы	33
7.1 Установка под дистрибутивы Linux	33
8. Удаление программы	37
8.1. Удаление с дистрибутивов Linux	37

1. Описание системы

DESKTOP программа SOICA Creator предназначена для использования в целях автоматизированного извлечения и редактирования полученной полезной информации из изображений, с последующим её применением в личных или рабочих целях.

За счёт нового инновационного подхода к обработке изображений, заключающегося в более глубоком понимании изображений и их элементной базы, SOICA Creator качественно распознаёт текст, таблицы, наличие печатей и подписей на изображениях с большим количеством шумов, высоким уровнем размытости и искажений. Программа имеет большой набор инструментов для редактирования распознанного текста и элемента, с возможностью добавления новых элементов, отсутствующих на исходном изображении (таких как таблица), что позволяет быстро скорректировать и получить желаемый результат до экспортирования документа.

SOICA Creator разворачивается на компьютере пользователя, работает автономно, доступ к интернету или браузеру не требуется.

Импортировать можно как одностраничные, так и многостраничные документы. Поддерживаемые форматы изображений: PDF, TIFF, JPEG, BMP, PNG. Содержание документов может быть любым. Классификация документов не предусматривается.

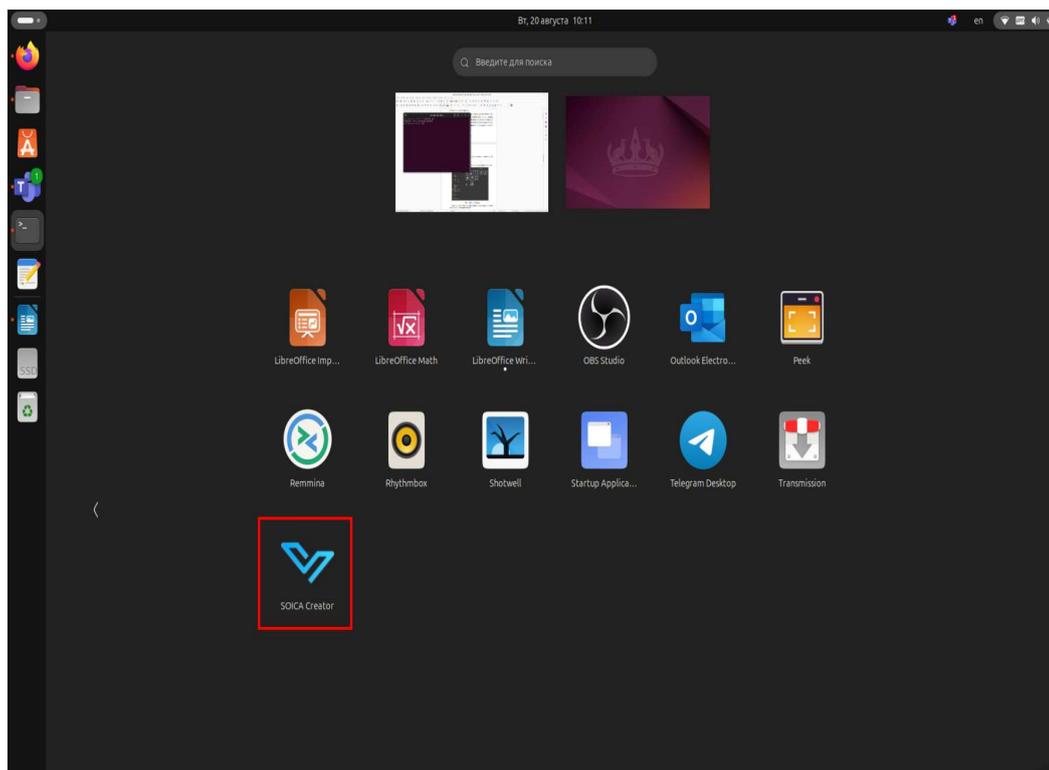
Экспортировать можно документы в формате ODT, PDF, PDF+T, PNG, JPEG, TIFF, JSON, содержащие распознанный текст, таблицы, подписи, печати и очищенный или оригинальный вариант изображения. Платформа не предусматривает комплектацию документов. Результаты распознавания будут привязаны к каждой странице с привязкой к координатам найденного элемента на каждой странице.

2. SOICA Creator

На данный момент платформа разворачивается локально на ПК пользователя.

2.1 Вход

Для запуска программы пользователю необходимо кликнуть по ярлыку.



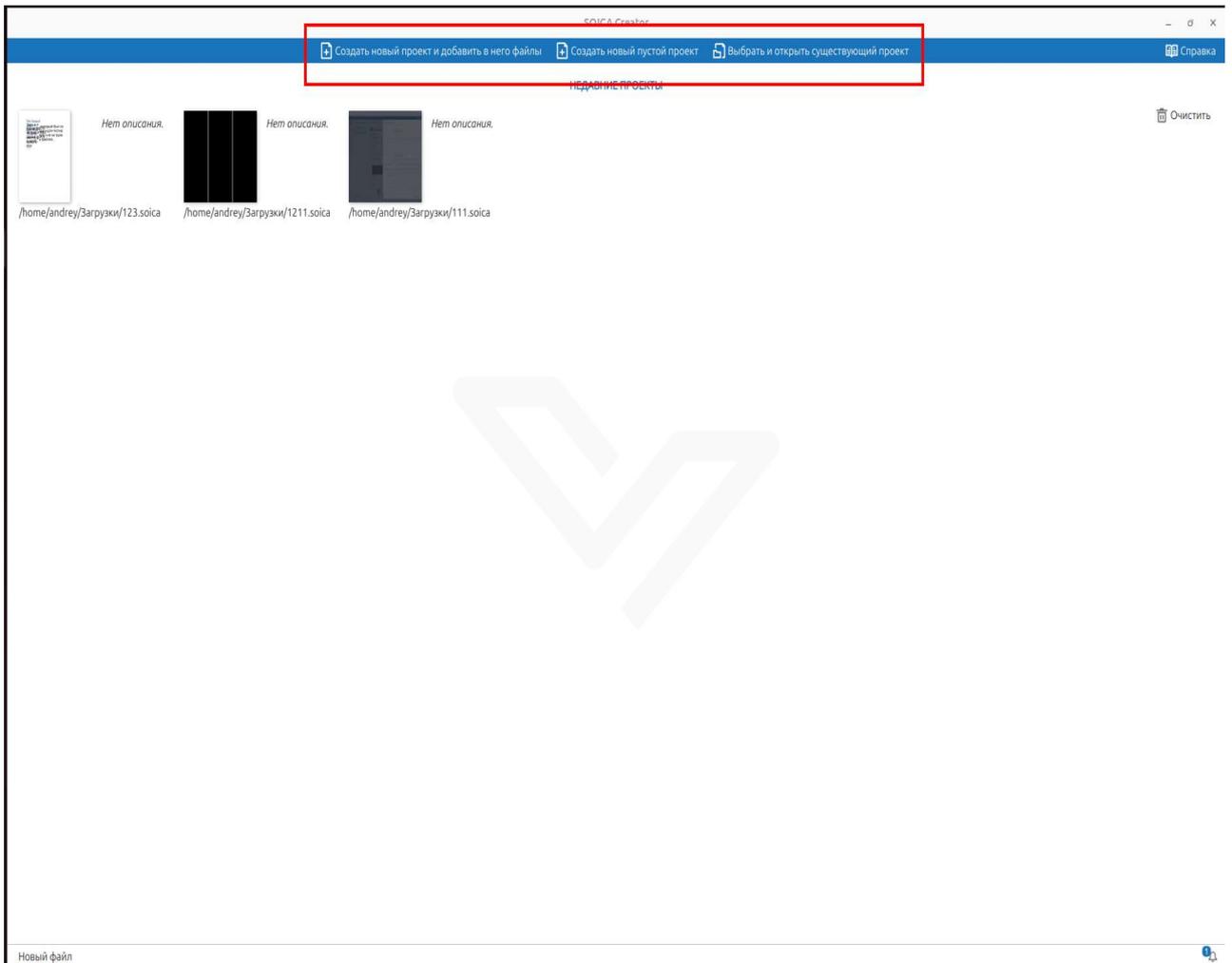
(Рис.1 Запуск программы)

2.2 Главный экран программы.

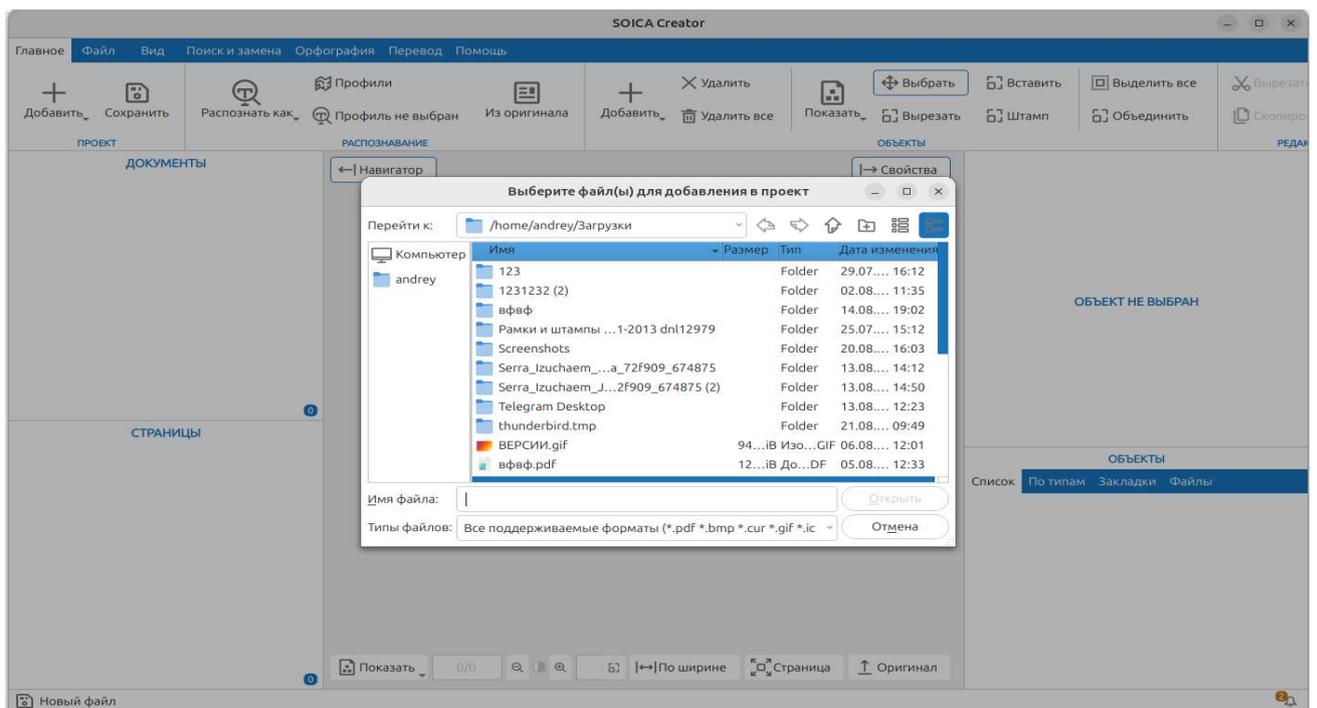
После завершения инициализации пользователь переходит на стартовую страницу.

В правом нижнем углу знак колокольчика  показывает системные уведомления (сообщение или ошибку). При появлении красного индикатора можно посмотреть их нажав на иконку.

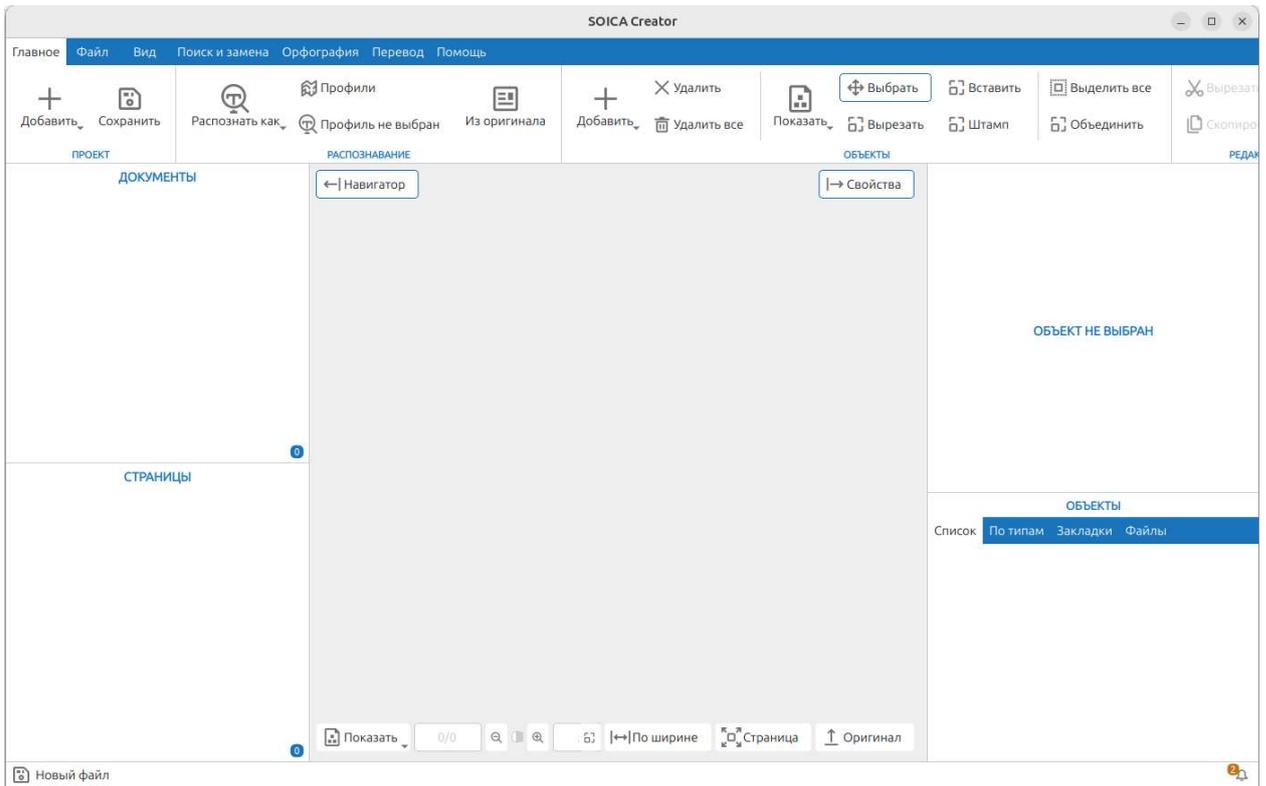
Наверху по центру расположено название программы SOICA Creator. Под ним расположены 3 кнопки с различными вариациями создания проекта.



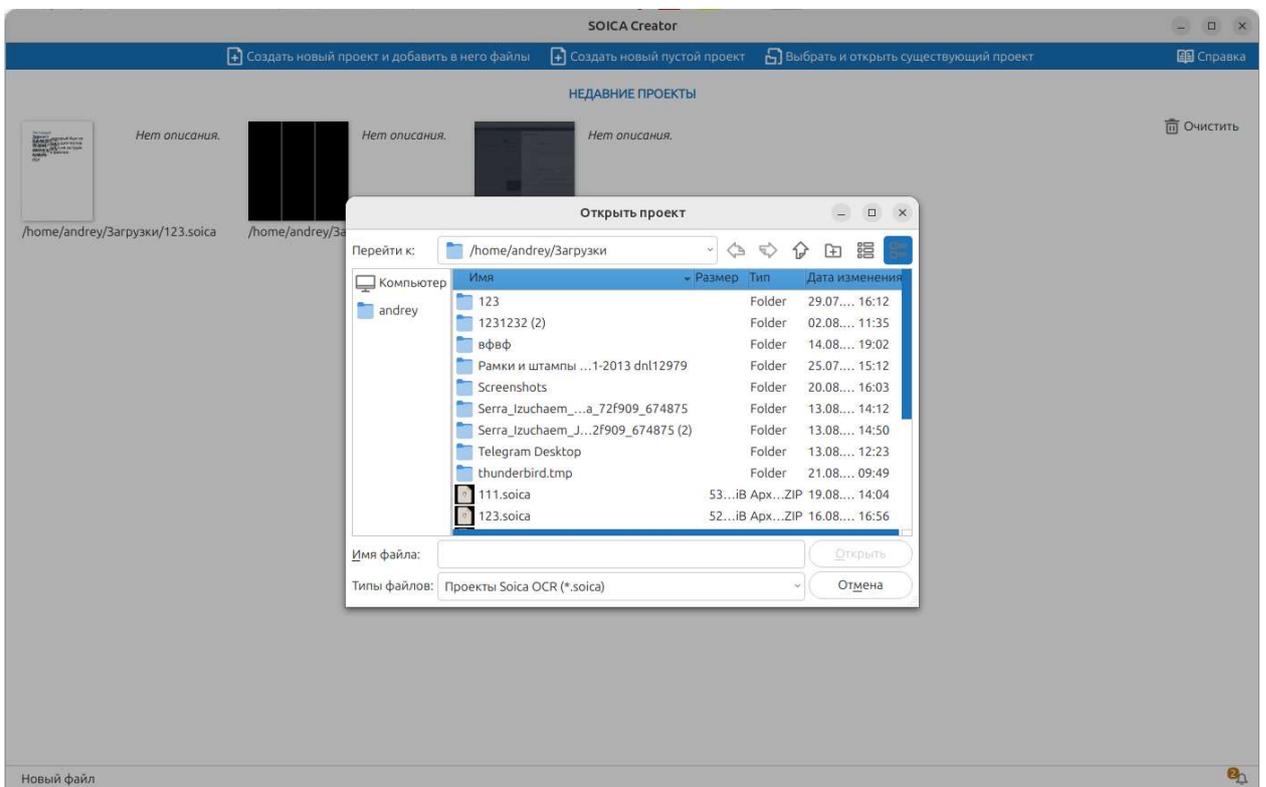
(Рис.2 Варианты создания проекта)



(Рис.3 Создание нового проекта и добавления в него файлы)

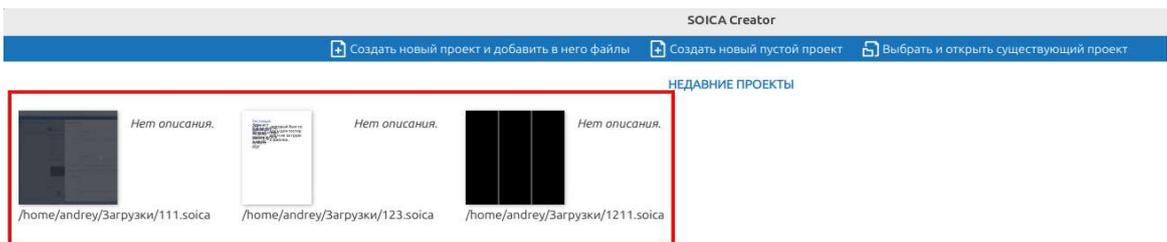


(Рис.4 Создание нового пустого проекта)



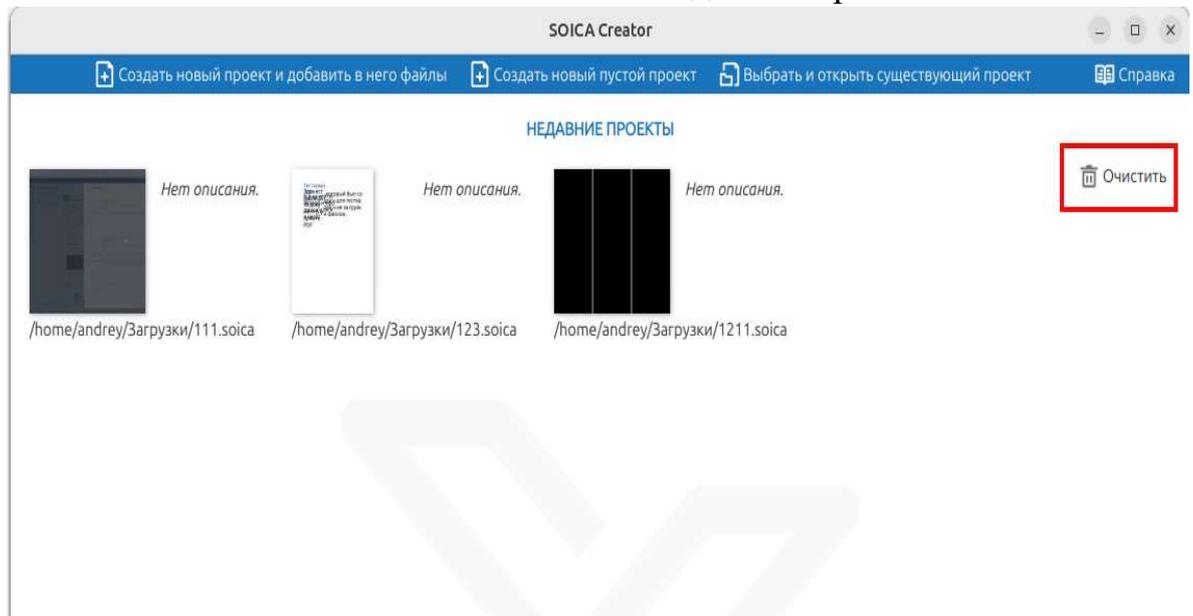
(Рис.5 Выбрать и открыть существующий проект)

Недавние проекты – позволяет открыть недавние проекты



(Рис.6 Открытие недавних проектов)

Очистить – позволяет очистить список недавних проектов.



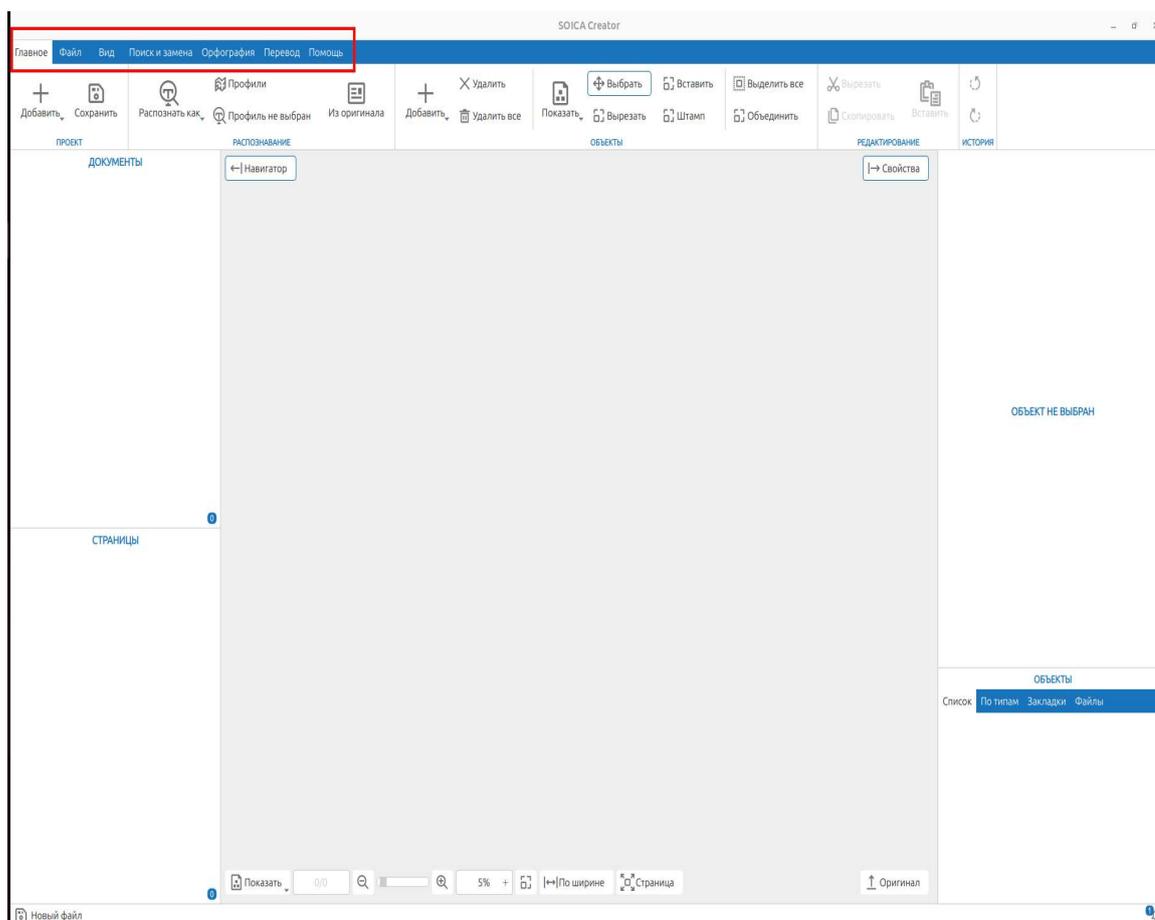
(Рис.7 Очистка списка недавних проектов)

Справка – позволяет получить доступ к подсказкам для программы.



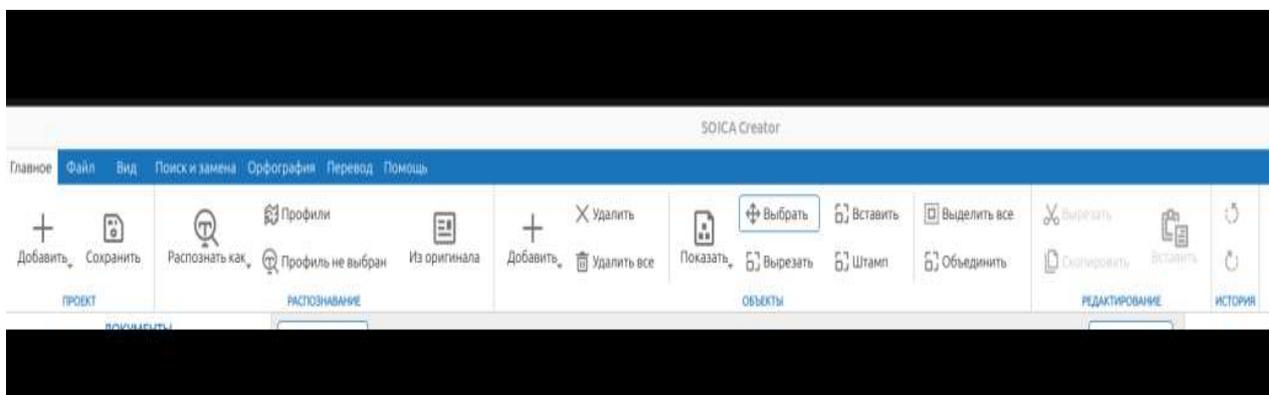
(Рис.8 Вызов справки)

После выбора одного из вариантов создания проекта или выбора одного из недавних проектов, происходит переход в меню программы SOICA Creator, где сверху расположено 7 кнопок меню с различным функционалом, рассмотрим каждую из них по отдельности.

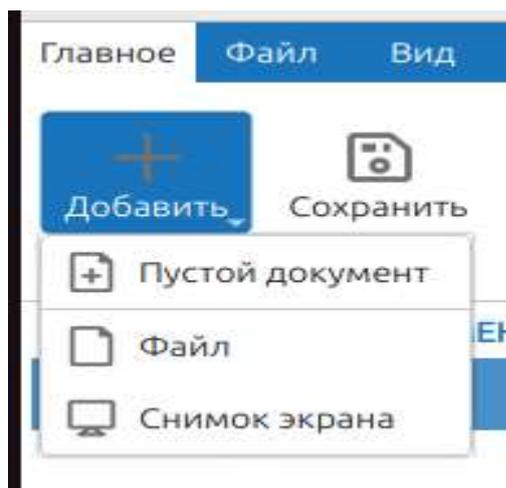


(Рис.9 Меню Soica Creator)

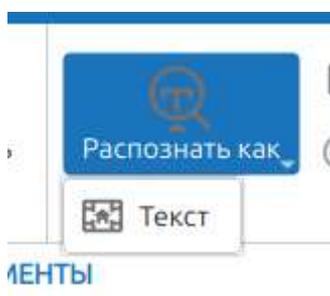
Главное – позволяет добавлять Файл или Объект, а, так же выполнять редактирование и распознавание над ним: Добавить/Сохранить файл, распознать как..., выбор профиля распознавания, редактирование объекта, история изменений.



(Рис.10 Меню Главное)

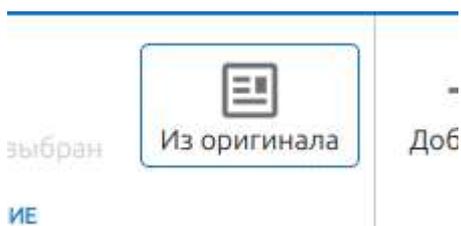


(Рис.11 Варианты добавления документа в проект)

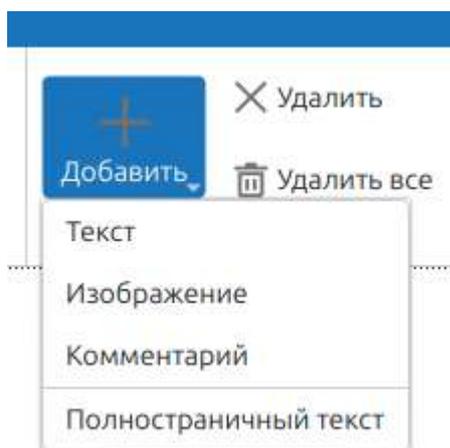


(Рис.12 Варианты распознавания)

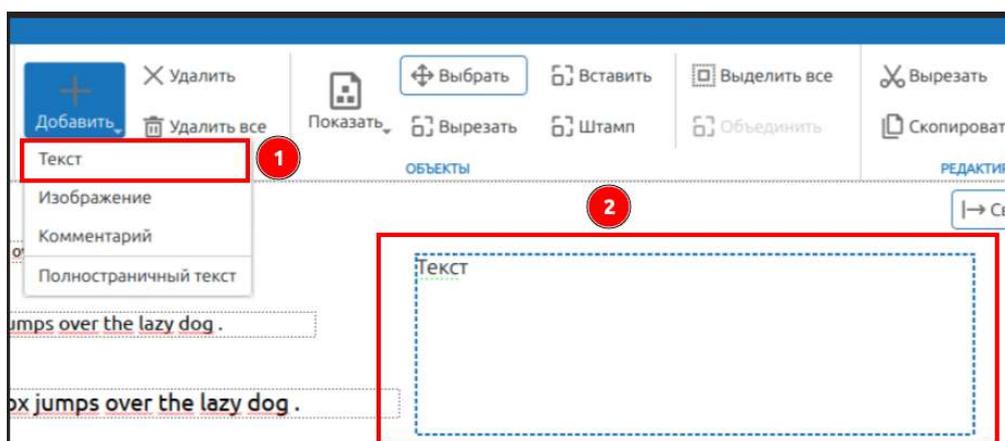
Из оригинала — инструмент позволяющий убрать все внесенные изменения с документа и оставляет оригинал документа.



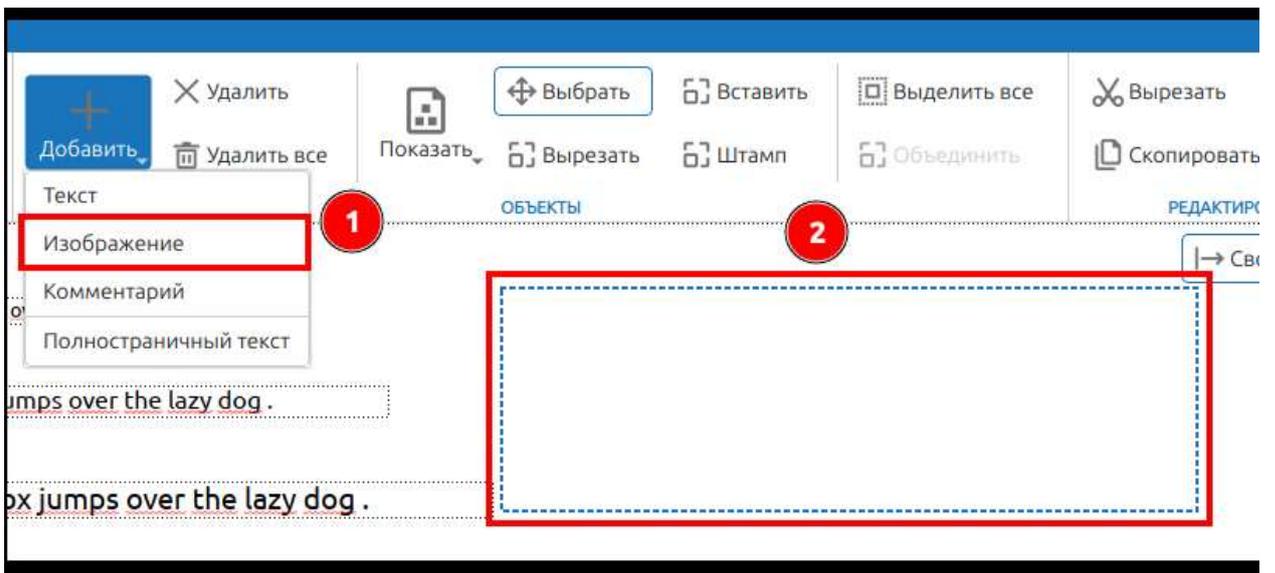
(Рис.13 Возможность вернуть документ в исходное положение)



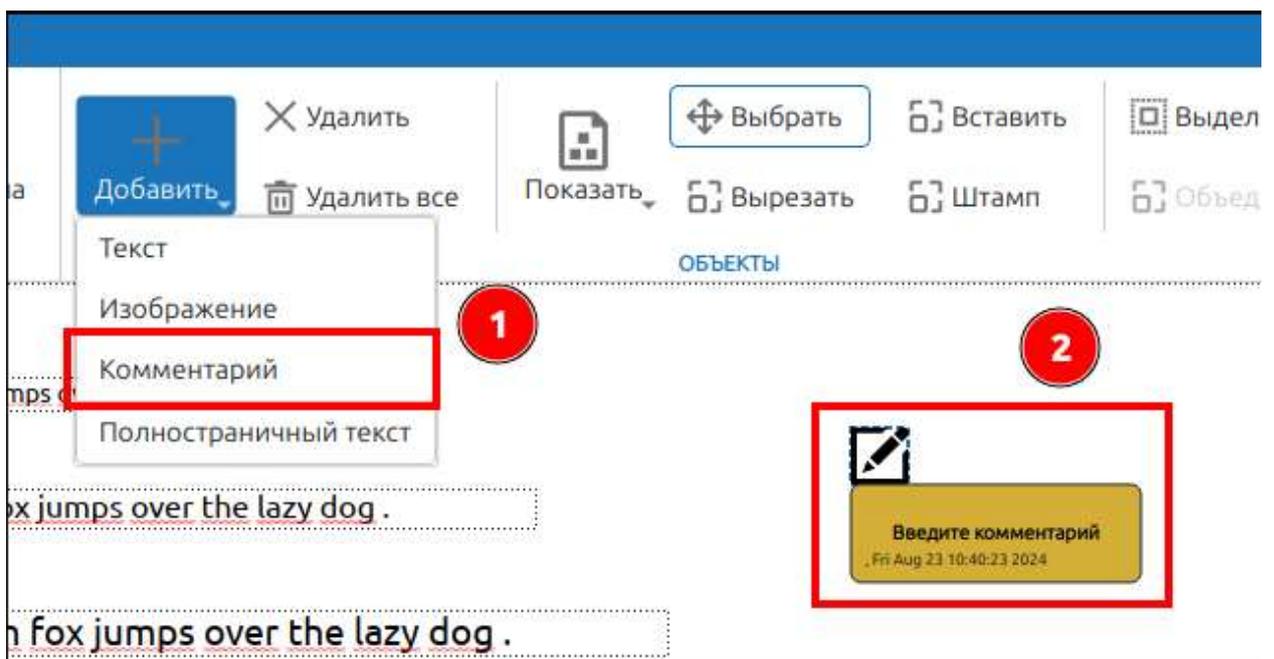
(Рис.14 Варианты добавления объекта)



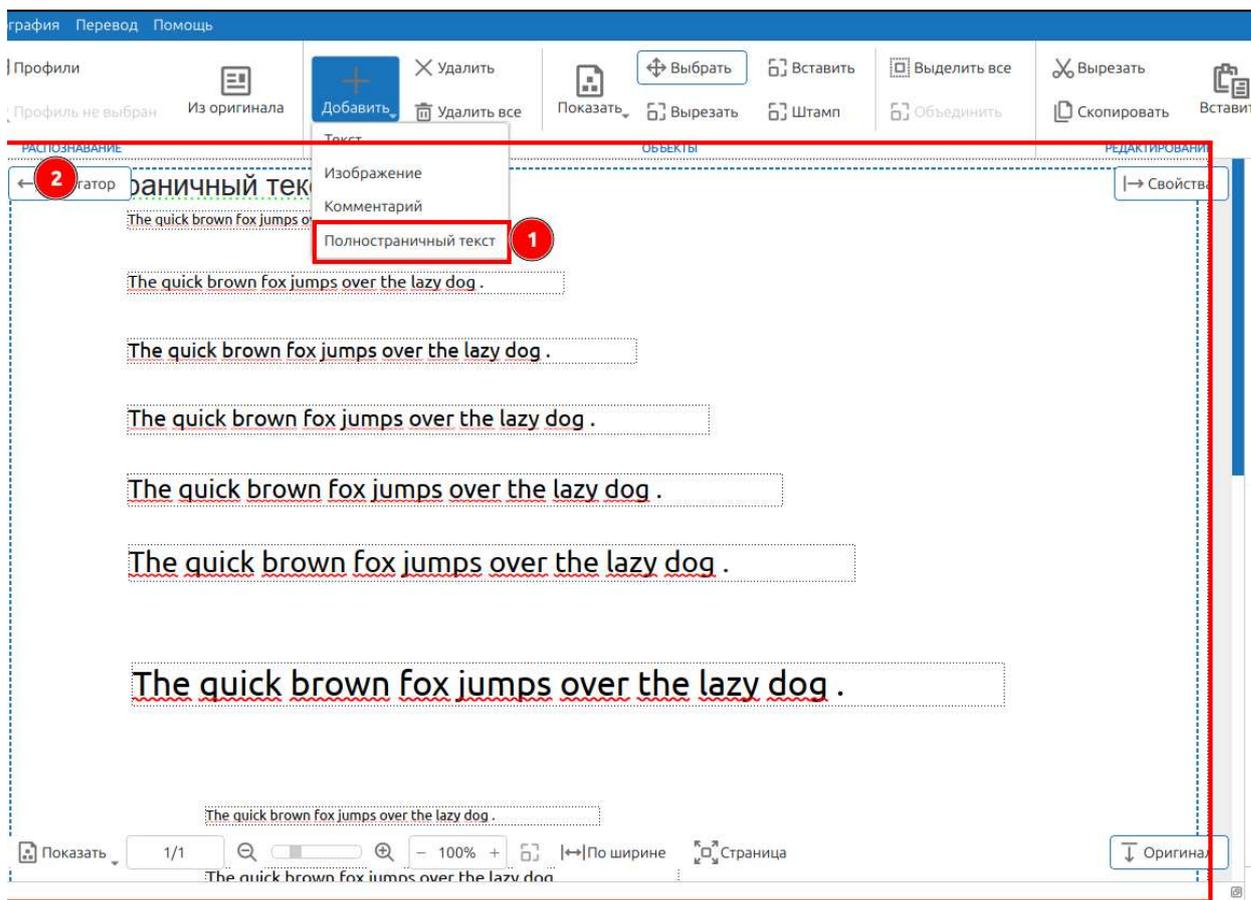
(Рис.15 Добавление объекта Текст)



(Рис.16 Добавление объекта Изображение)

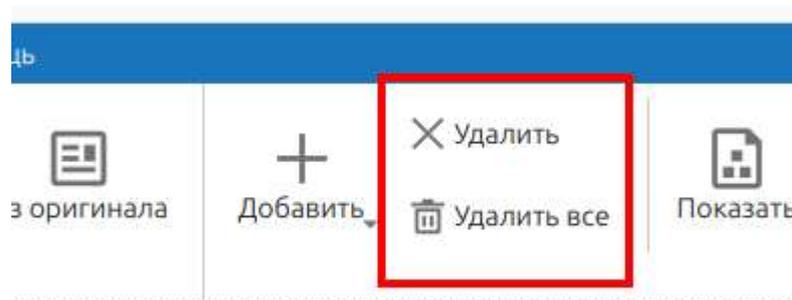


(Рис.17 Добавление объекта Комментарий)



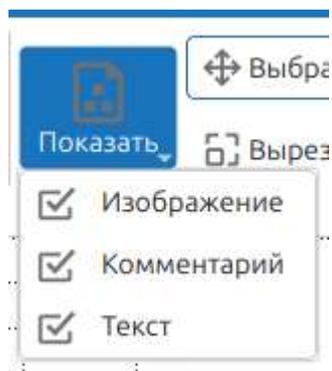
(Рис.18 Добавление объекта Полностраничный текст)

Для удаления одного, нескольких (необходимо выделить несколько) или всех объектов есть функции удаления.



(Рис.19 Удаление одного, нескольких или всех объектов)

Для отображения нужных элементов необходимо нажать Показать, выбрать и поставить галочку в чек боксе рядом с нужными из списка элементами.



(Рис.20 Варианты отображение объектов в документе)

Есть панель инструментов для объектов: выбрать, вставить, вырезать, добавить штамп, выделить все.



(Рис.21 Панель инструментов для объектов)

Выбрать — инструмент для выбора и перемещения объектов.

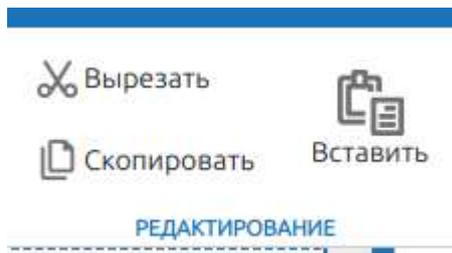
Вырезать — инструмент для выделения части страницы и перемещение в буффер обмена.

Вставить — инструмент для вставки изображения из буффер обмена.

Штамп — инструмент добавления штампа.

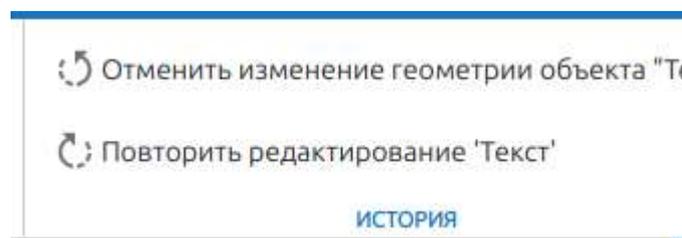
Выделить все — инструмент для выделения всех объектов.

Есть панель редактирования позволяющая вырезать, скопировать, вставить выделенный элемент на страницы документа.



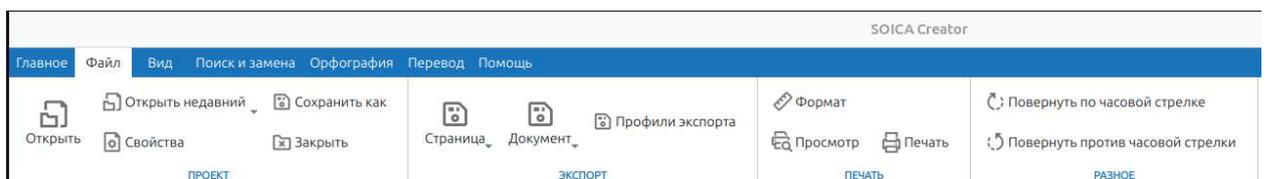
(Рис.22 Панель редактирования.)

Есть панель управления историей страницы. На ней можно управлять изменениями сделанных на странице документа: отменить или повторить действие.



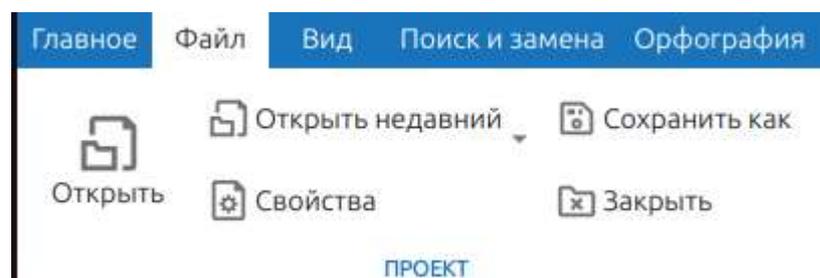
(Рис.23 Панель управления историей.)

Файл — позволяет открывать недавний или созданный ранее проект, а так же доступна возможность экспорта и печати. Позволяет менять ориентацию документа.



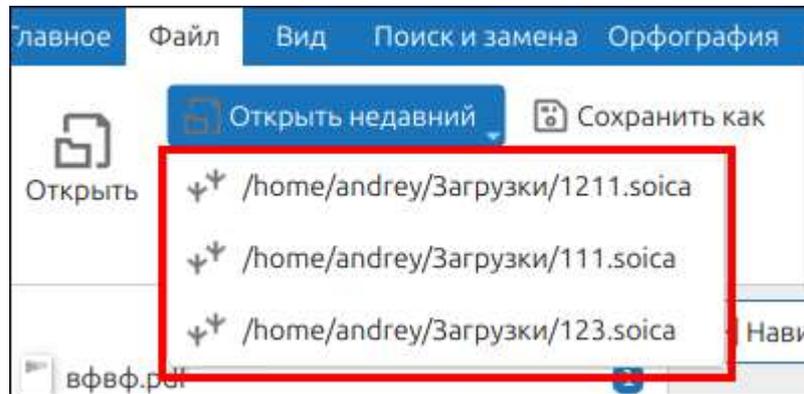
(Рис.24 Меню Файл)

В панели управления проектом можно выполнить следующие действия: открыть или открыть недавний проект, сохранить проект, изменить свойства проекта и закрыть проект.



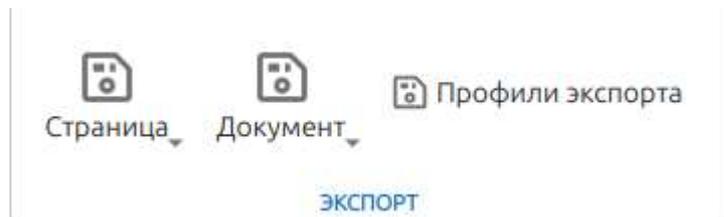
(Рис.25 Панель управления проектами.)

Открыть недавний — позволяет открыть проект, в котором ранее проводилась работа.

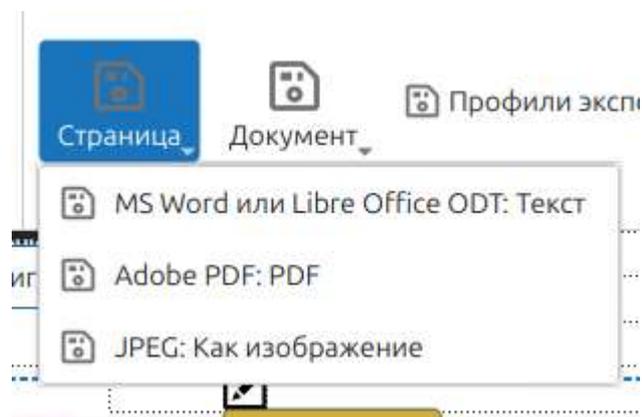


(Рис.26 Открытие недавних проектов)

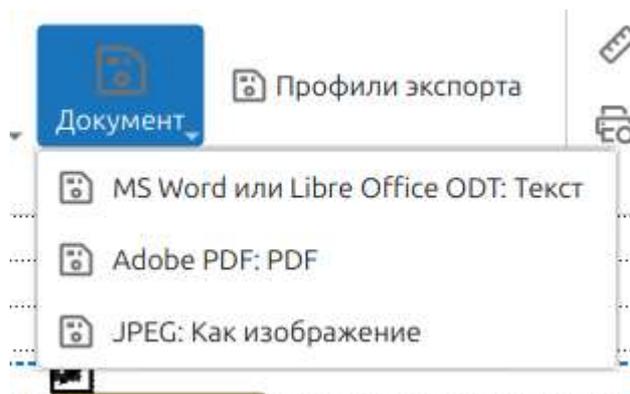
Есть возможность экспорта страниц и документа.



(Рис.27 Экспорт страниц и документа)

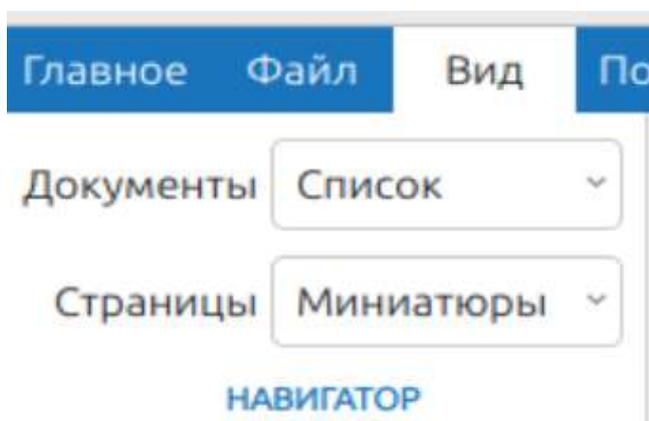


(Рис.28 Варианты экспорта страницы)



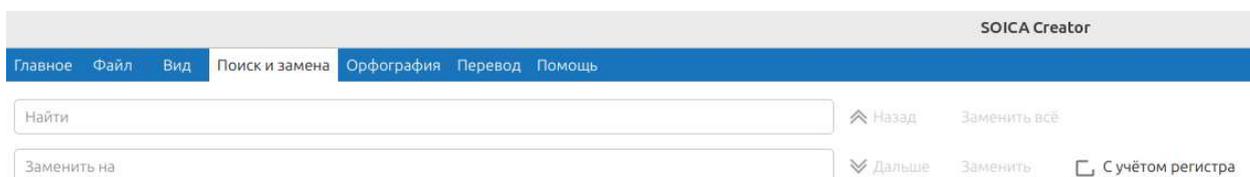
(Рис.29 Варианты экспорта документа)

Вид – позволяет отобразить документы и страницы в виде списка и миниатюры.



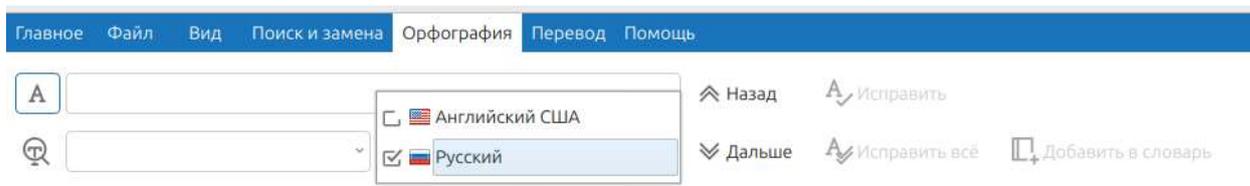
(Рис.30 Меню Вид)

Поиск и замена – позволяет найти часть текста и заменить его.



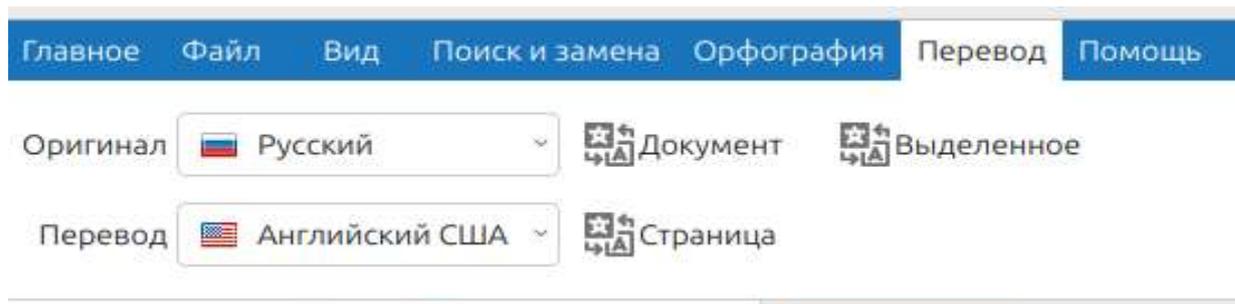
(Рис.31 Меню Поиск и замена)

Орфография – позволяет выполнить проверку орфографии и поиск неточно распознанных слов.



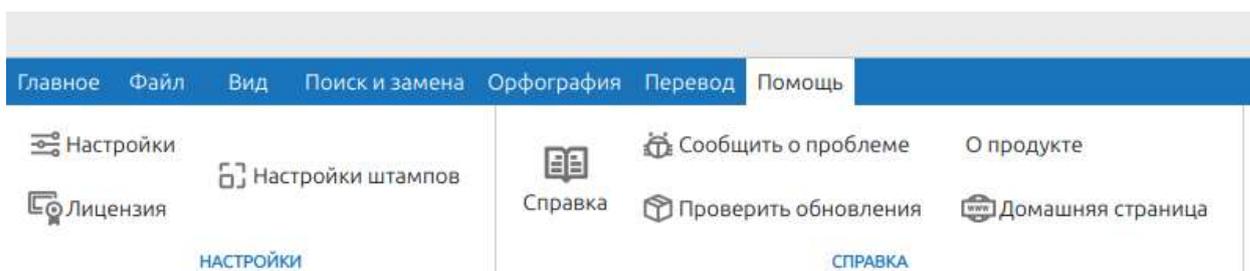
(Рис.32 Меню Орфография)

Перевод – позволяет перевести текст.

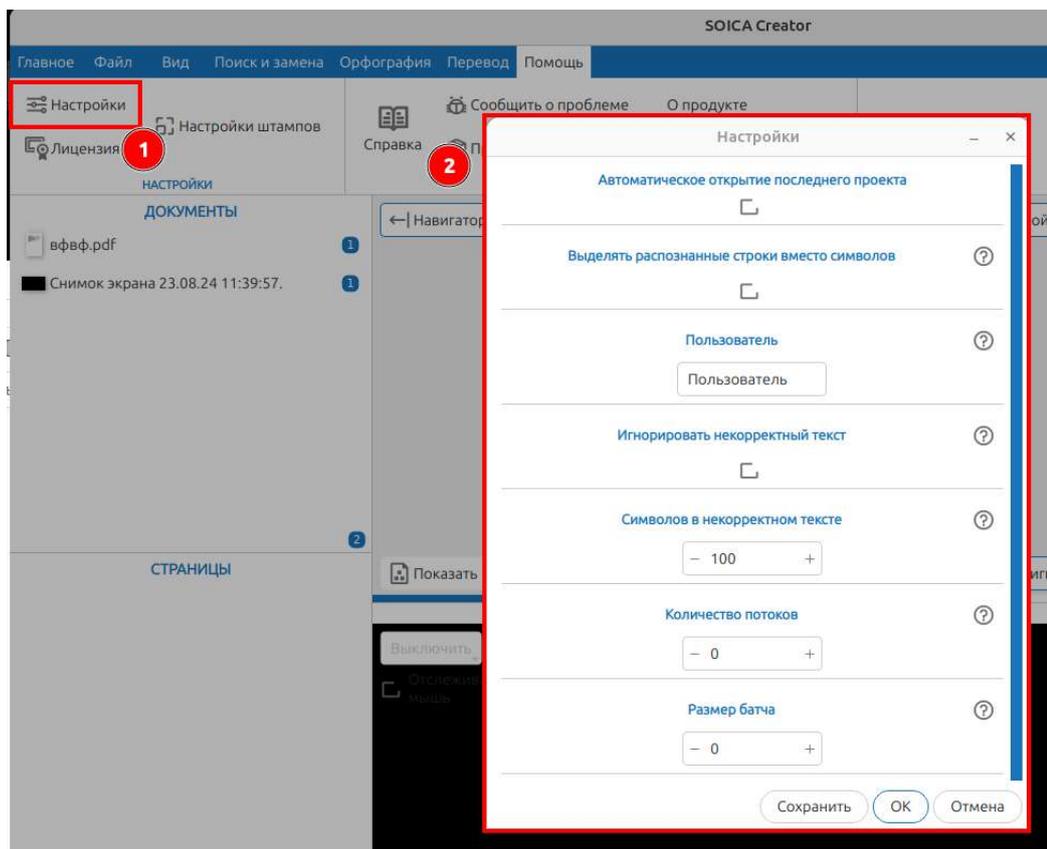


(Рис.33 Меню Перевод)

Помощь – позволяет взаимодействовать с настройками программы, настраивать штампы. Предоставляет доступ к просмотру лицензии, справке, информации о продукте. Возможность проверки обновления. Переход на сайт проекта. Обратная связь.

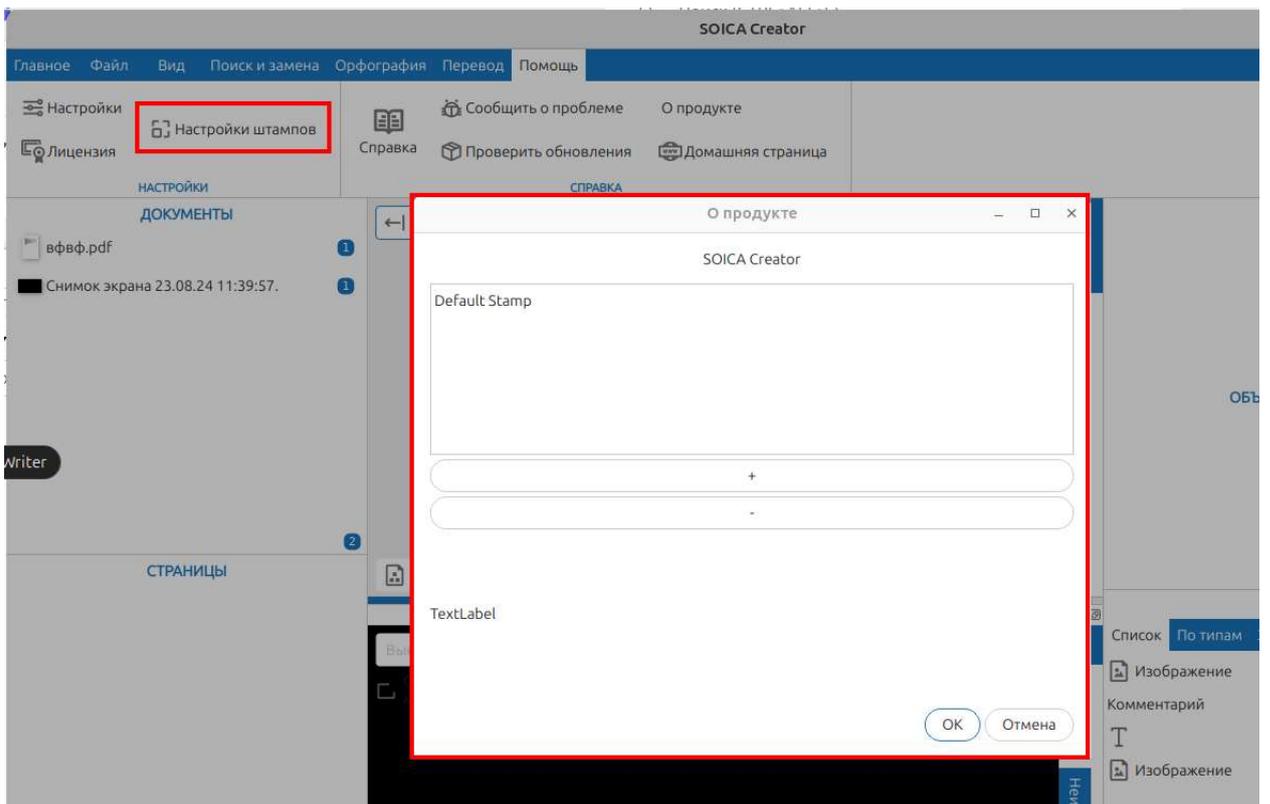


(Рис.34 Меню Помощь)



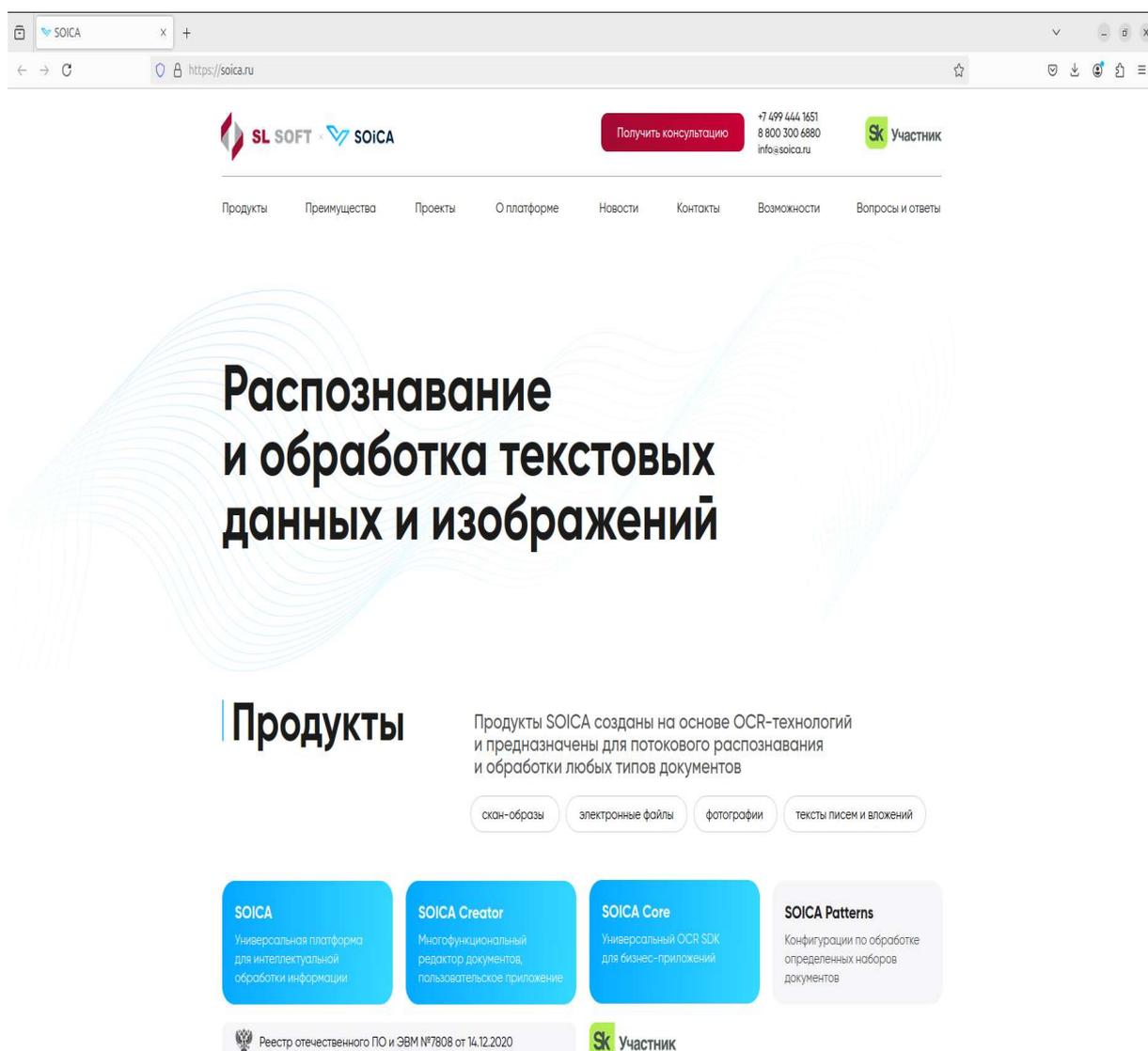
(Рис.35 Настройки программы)

Настройки штампов — позволяет добавить новые штампы, а так же настроить их.



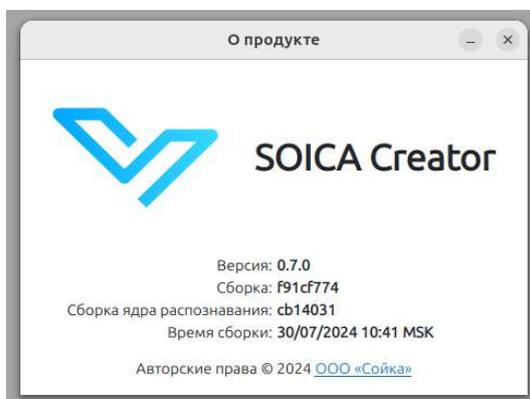
(Рис.36 Настройки штампов)

При выборе «Домашняя страница» осуществляется переход на официальный сайт Soica.ru, на котором можно ознакомиться со всеми возможностями платформы Soica и заказать звонок.



(Рис.37 Домашняя страница. Официальный сайт Soica.ru)

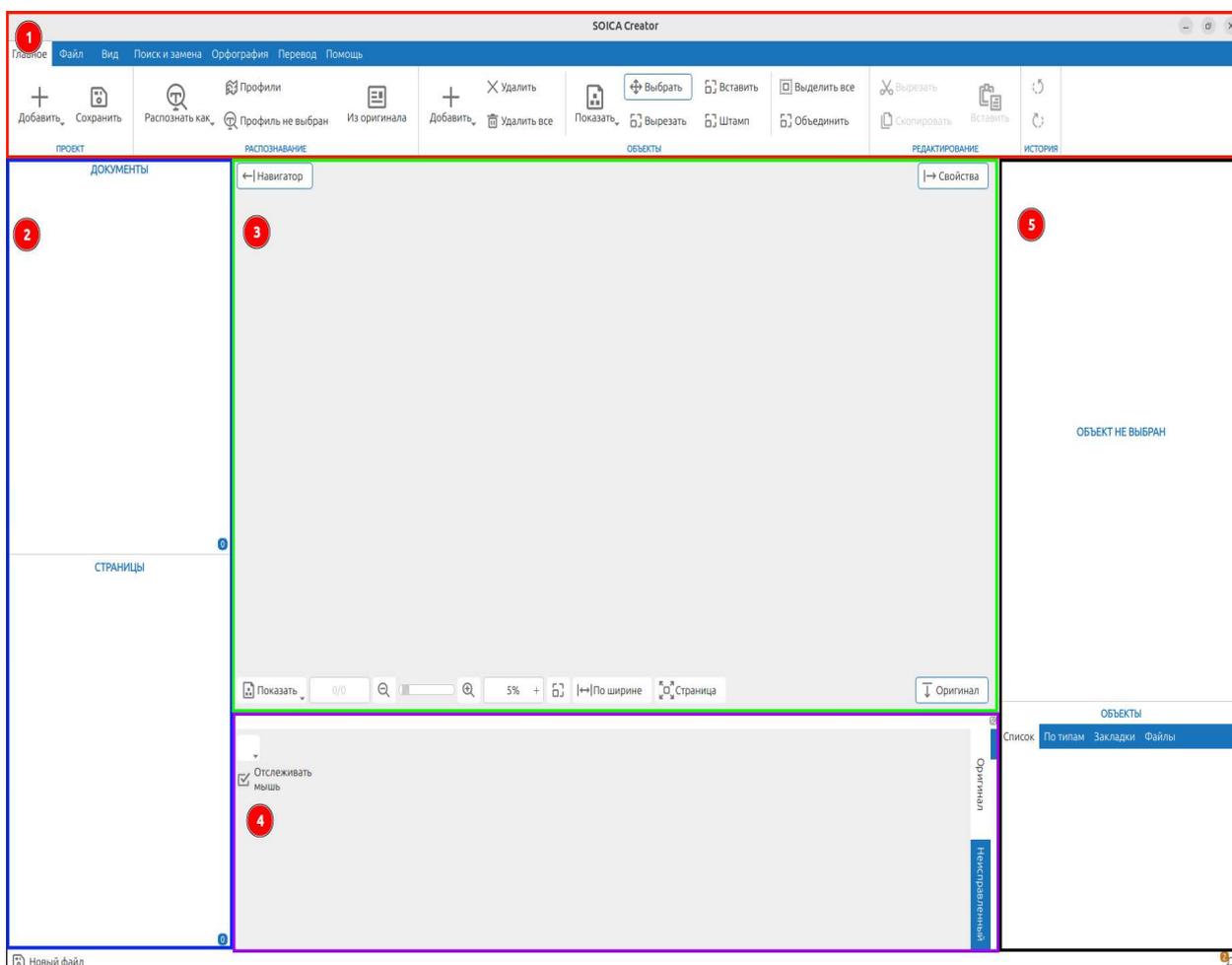
При выборе «О продукте» можно увидеть подробную информацию об установленной версии системы.



(Рис.38 О продукте. Информация об установленной версии системы)

3 Интерфейс проекта

После того, как пользователь создал новый проект интерфейс системы меняется и состоит из 5 зон:



(Рис.39 Общий вид интерфейса нового проекта)

Зона 1. Панель переключения между меню. Возможности Панели общей настройки рассмотрены выше, в п. 2.2. Главный экран программы.

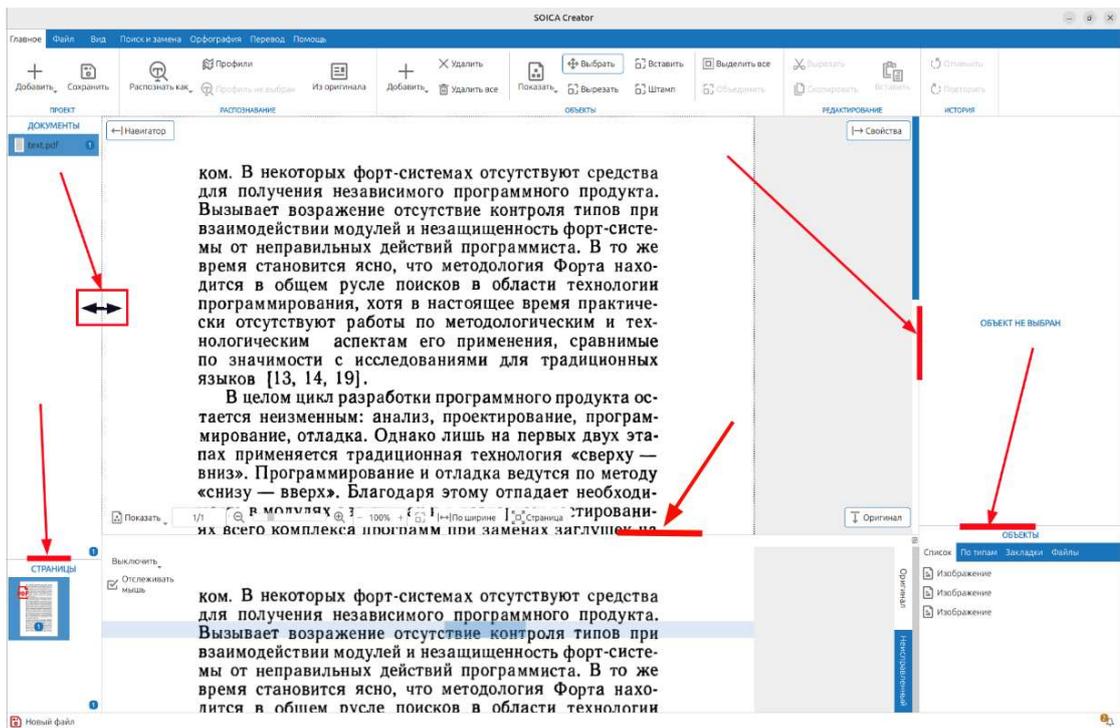
Зона 2. Навигационное меню.

Зона 3. Панель работы со страницей документа.

Зона 4. Панель работы с оригиналом.

Зона 5. Панель работы со свойствами объекта в документе.

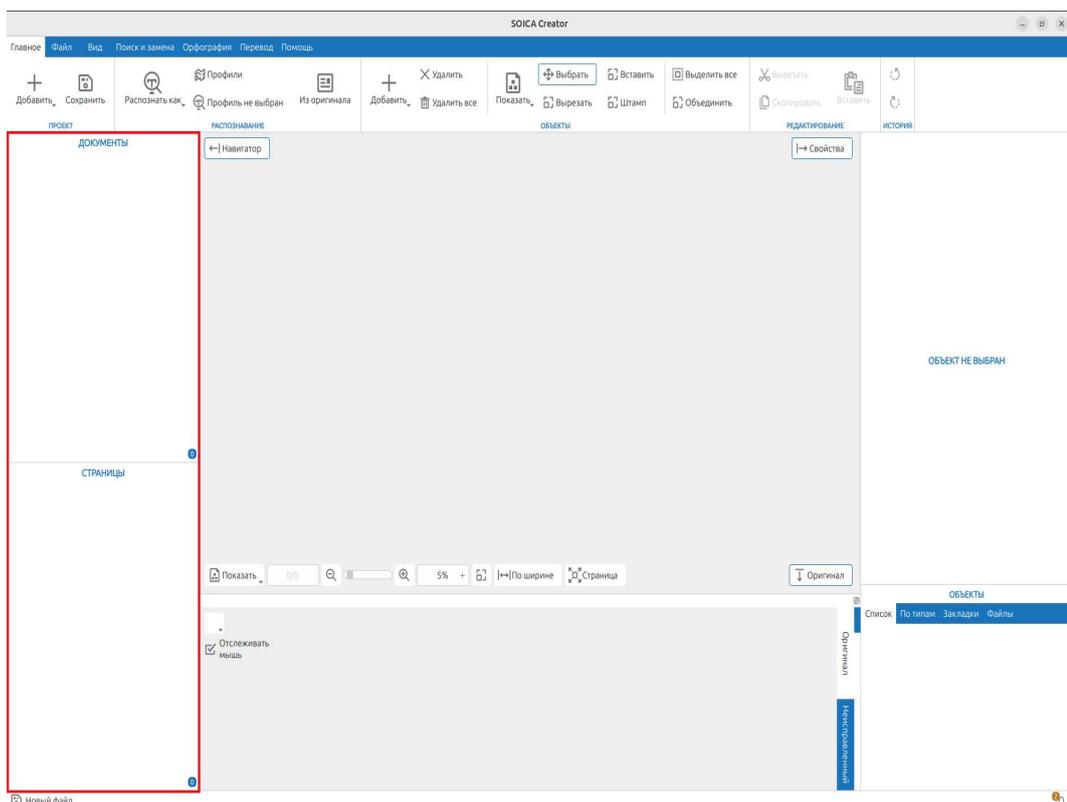
Размер зон 2, 4 и 5 можно изменить, потянув за край области, как показано на рисунке ниже.



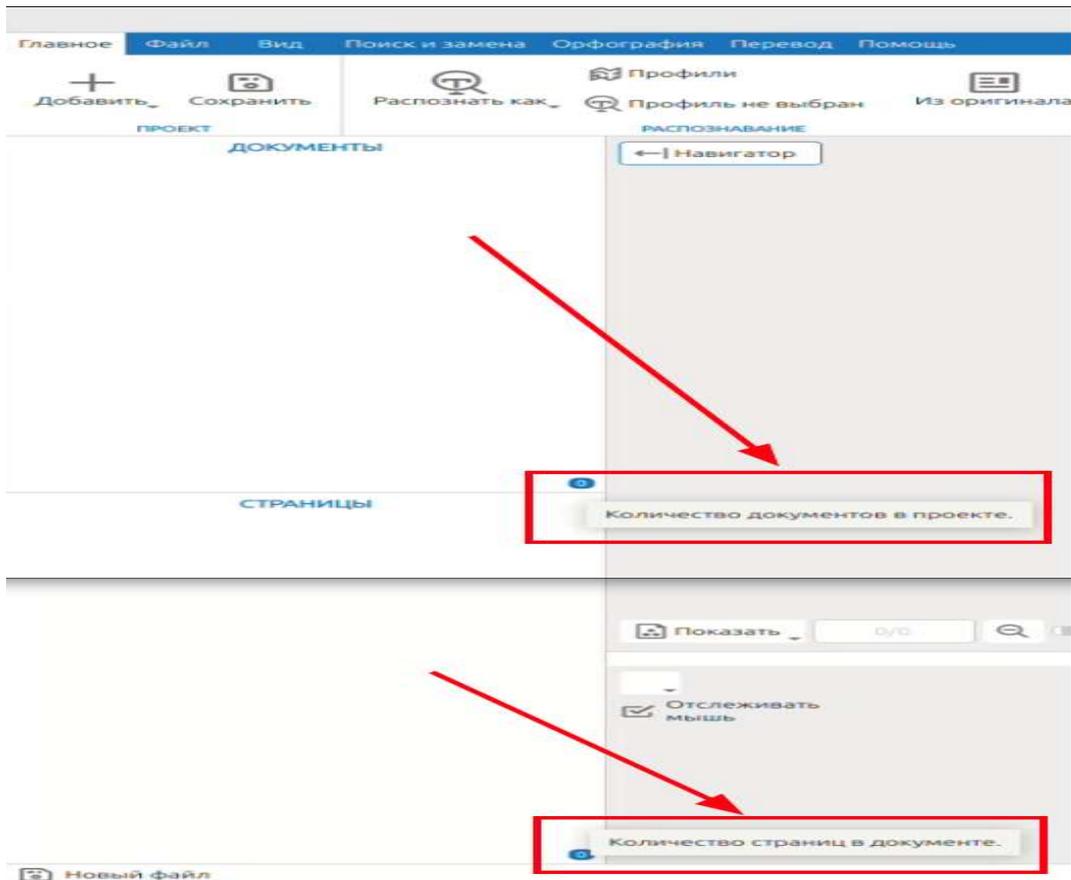
(Рис.40 Способ изменения размера области зон 2, 4 и 5)

3.1 Навигационное меню

Служит для работы с загружаемыми документами и их страницами. Делится на две части: Документы и Страницы.

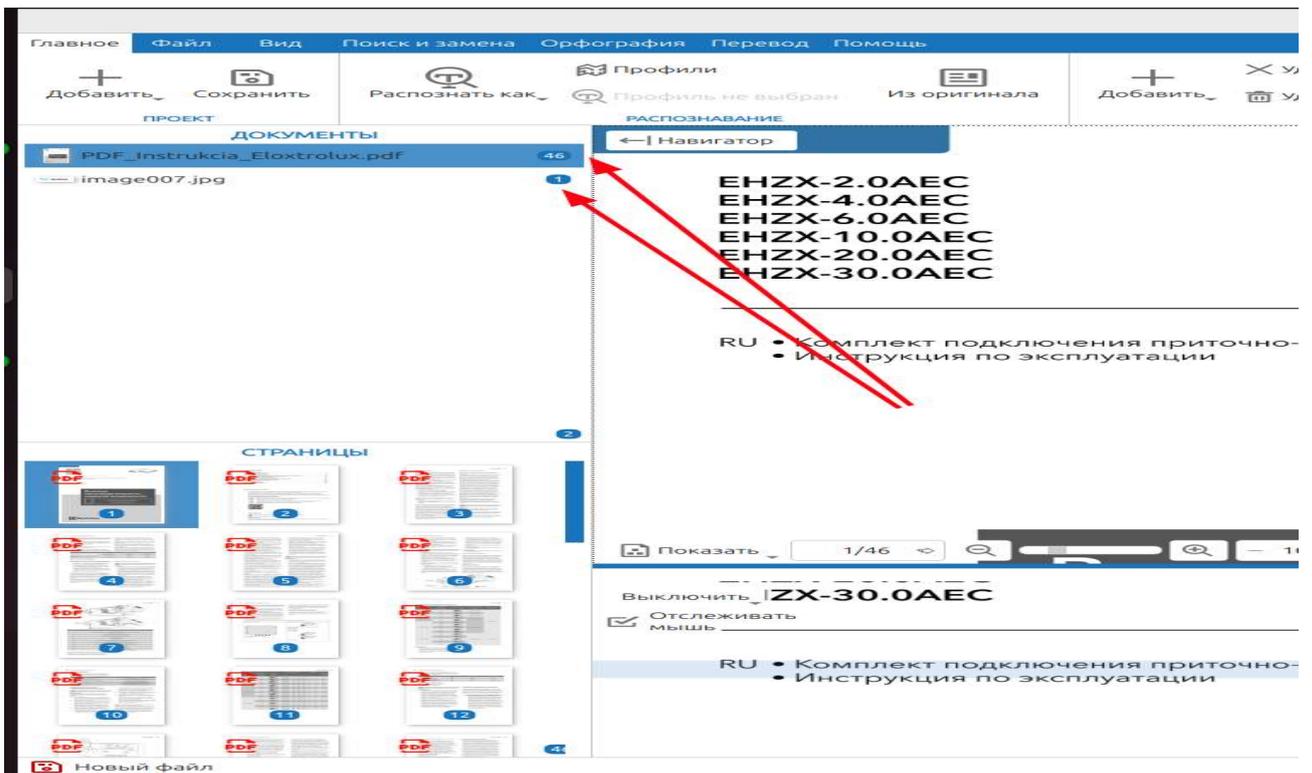


(Рис.41 Навигационное меню)



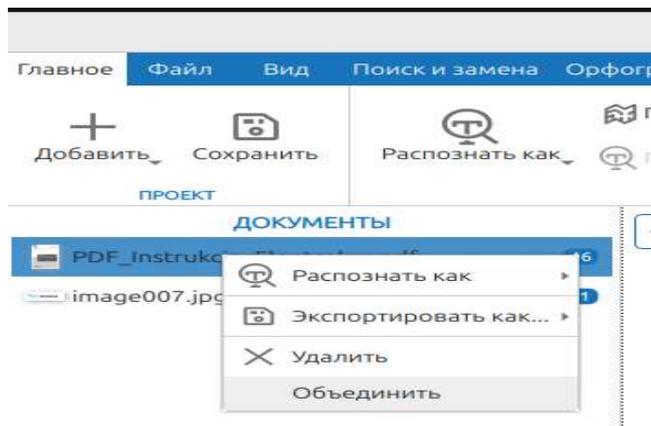
(Рис.42 Количество документов и страниц в проекте)

После добавления файлов в проект, изображения отображаются во всех зонах. В разделе документы можно увидеть название файла и количество добавленных документов. Рядом с отображаемым документом есть счетчик страниц.



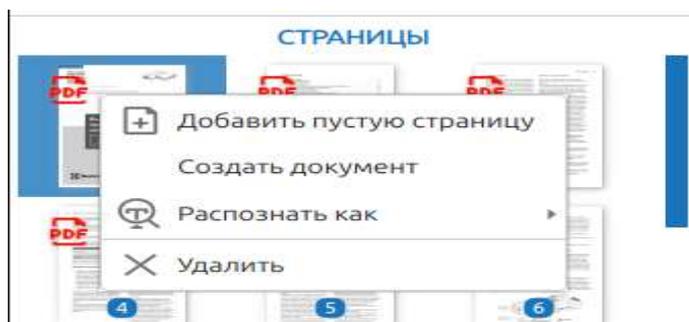
(Рис.43 Счетчик страниц документа)

Над документом можно выполнить следующие действия: распознать, экспортировать, удалить и объединить (если выделено 2 и более документа). Для этого необходимо выбрать документ и нажать правую кнопку мыши. Сценарии для распознавания и экспортирования необходимо настроить в разделе Главное и Файл.



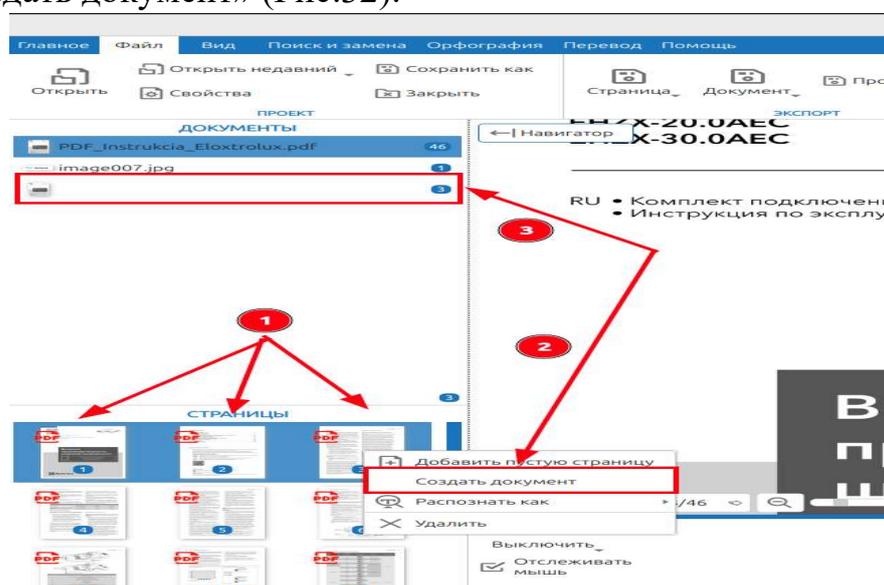
(Рис.44 Возможные действия над документом)

Над страницами можно выполнить следующие действия: добавить пустую страницу, создать документ, распознать (одну или несколько) и удалить (Рис.31). Сценарий распознавания необходимо настроить в разделе Инструменты.



(Рис.45 Возможные действия над страницами)

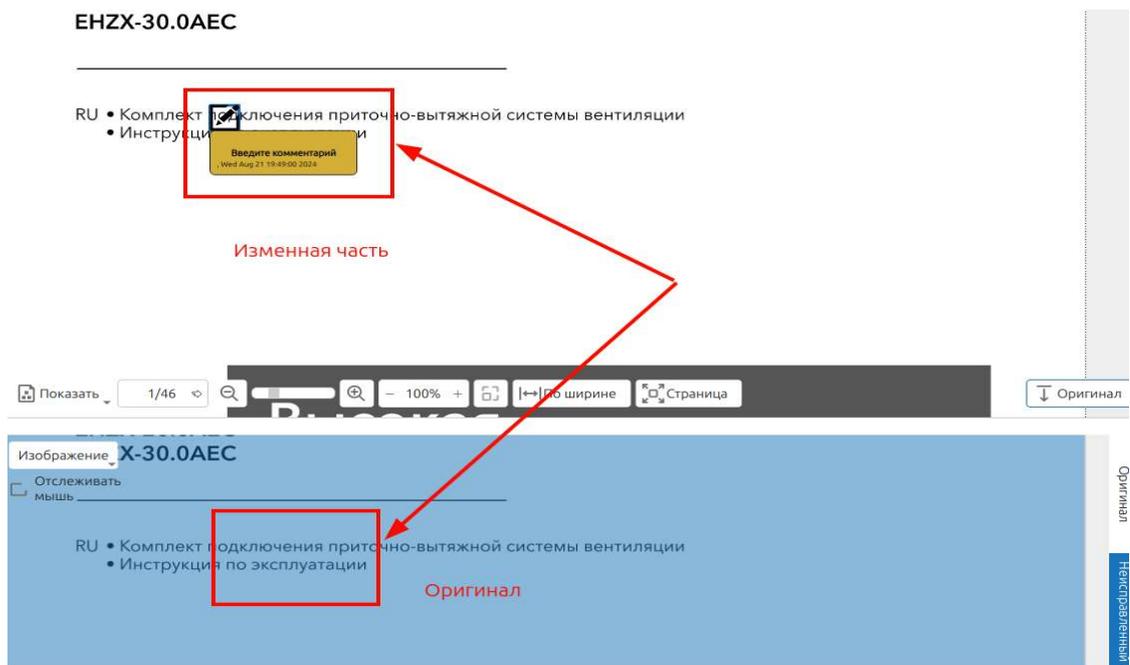
Есть возможность выделить нужные страницы в новый документ. Для этого необходимо в разделе «Страницы» выделить нужные страницы и выбрать «Создать документ» (Рис.32).



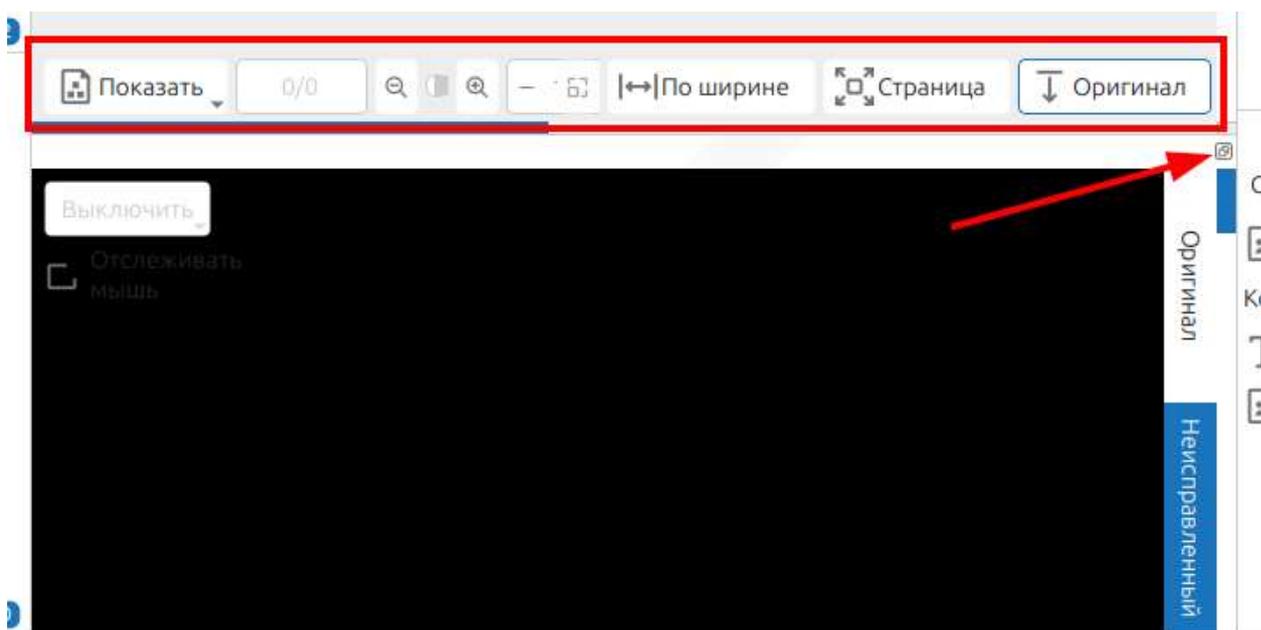
(Рис.46 Создание нового документа из выбранных страниц)

3.2 Панель работы со страницей документа.

В центральной части экрана (Рис.33) расположена панель работы со страницей документа, в которой можно удалить выделенные элементы, а так же скопировать, вырезать и вставить объекты. Скопировать и вставить стиль объекта.

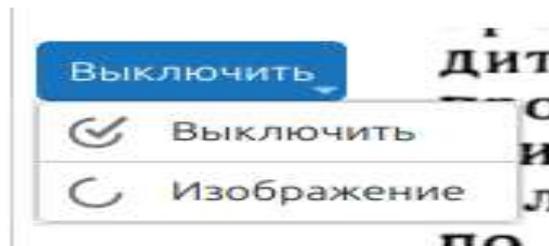


(Рис.49 Отслеживание изменений с панелью работы со страницей документа)



(Рис.50 Дополнительная панель управления для документа и оригинала)

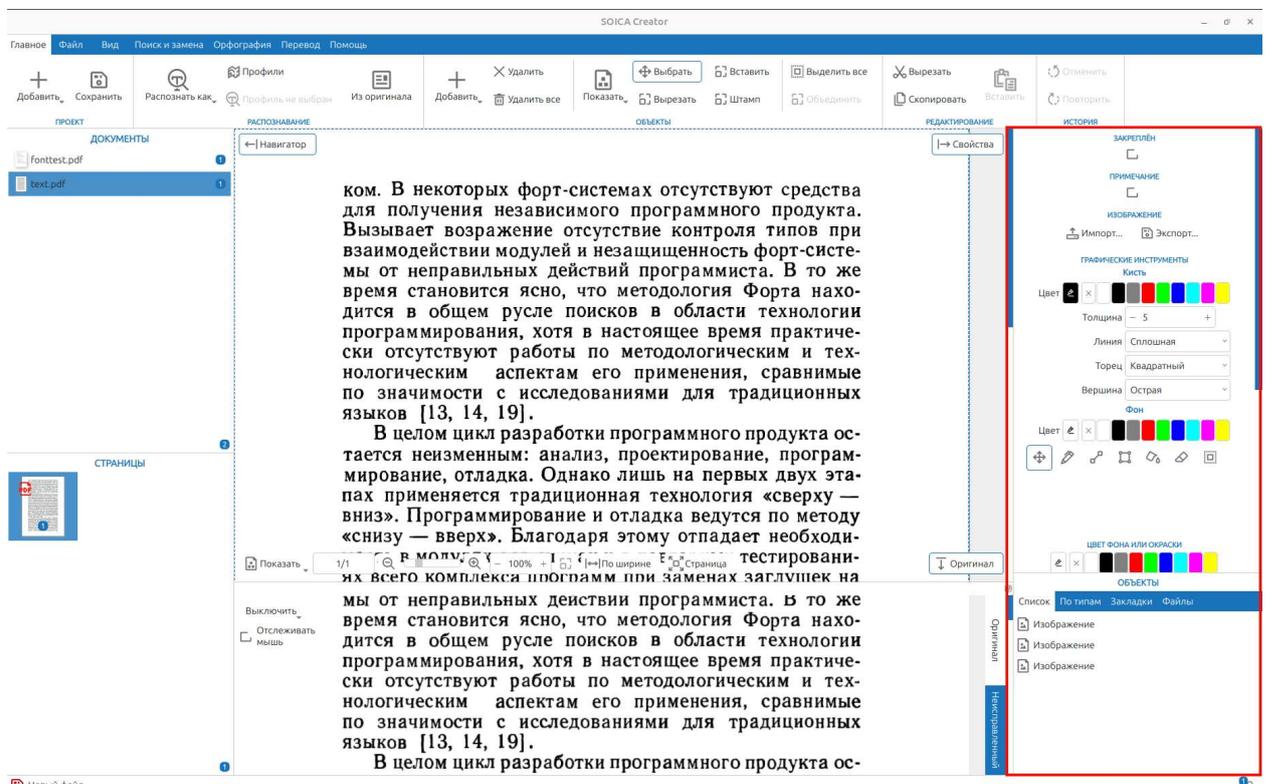
Можно указать настройки сегментации: Выключить, Изображение. Необходимо выбрать радиобаттоном.



(Рис.51 Настройка сегментации)

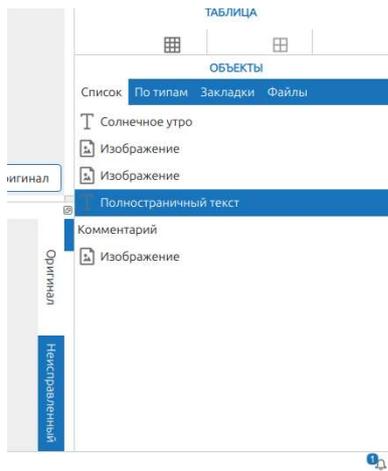
3.4 Панель работы со свойствами объекта в документе.

Панель работы со свойствами объекта в документе — набор инструментов, позволяющие редактировать объекты.

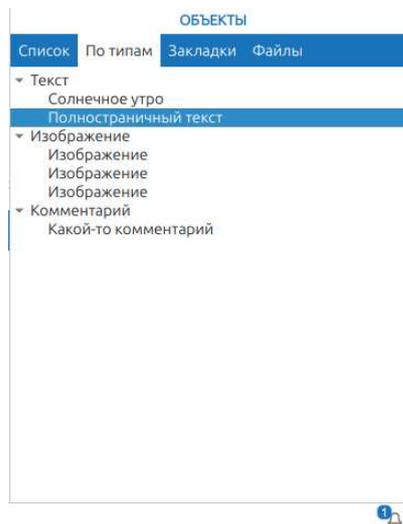


(Рис.52 Панель работы со свойствами объекта в документе)

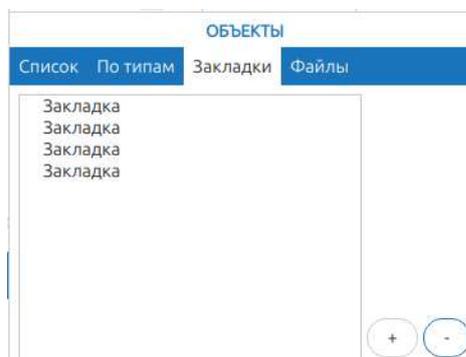
На панели работы со свойствами объекта в документе, есть возможность просмотреть список добавленных объектов, перейдя на вкладку Список. Можно сортировать этот список во вкладке По типам. Так же панель позволяет создать закладки по страницам и загрузить файлы.



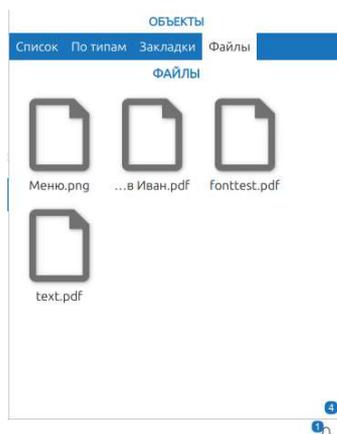
(Рис.53 Вкладка список)



(Рис.54 Вкладка По типам)



(Рис.55 Вкладка Закладки)

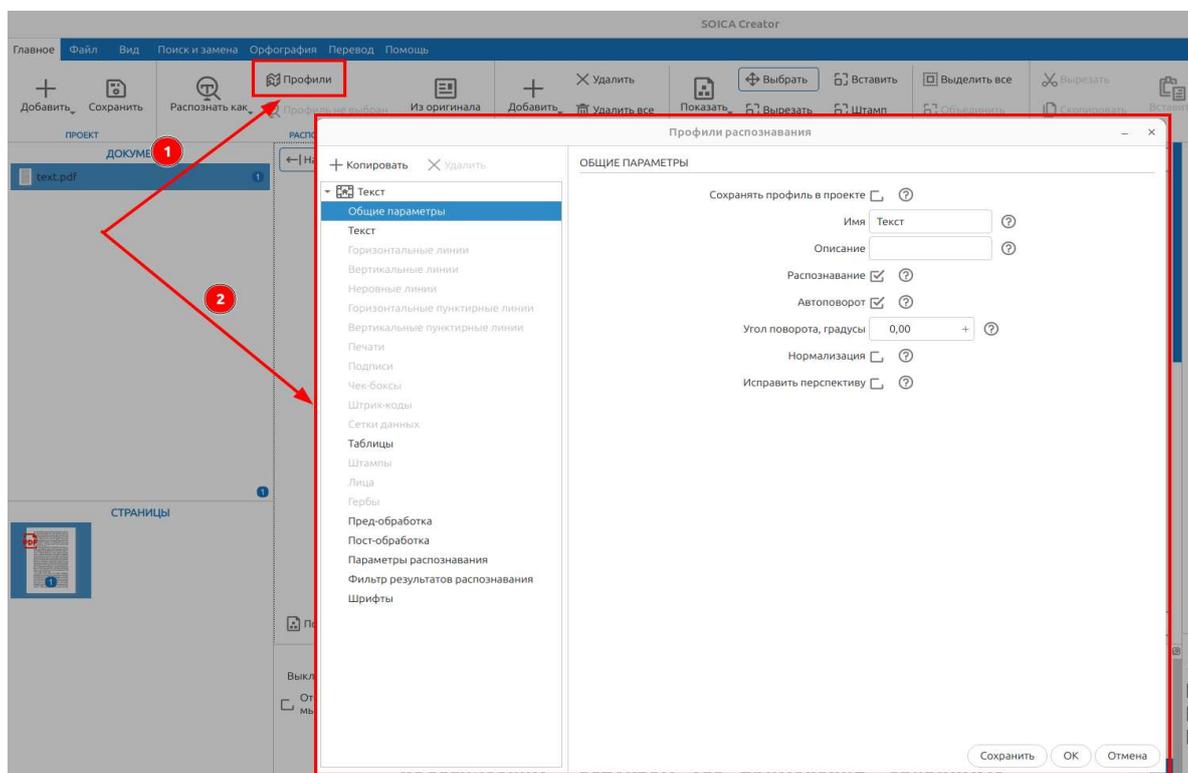


(Рис.56 Вкладка Файлы)

4. Профили распознавания.

Профиль распознавания – набор конфигурационных настроек, которые будут применяться при распознавании изображения выбранным профилем.

Чтобы создать, настроить или изменить профиль распознавания необходимо перейти в меню Главное, далее в Профили. Что бы изменить параметры, необходимо перейти в общие параметры профиля распознавания.



(Рис.57 Переход к настройкам профилей распознавания)

Можно копировать новый профиль распознавания с помощью кнопки Копировать или удалить уже существующий профиль распознавания.

Сохранить профиль в настройках приложения. Если этот параметр включен, профиль экспорта будет сохранён в настройках приложения. Иначе он будет сохранён в проекте.

Имя. Уникальное название профиля экспорта.

Описание. Комментарии к профилю экспорта.

Распознавание. Извлекать текст из графического изображения.

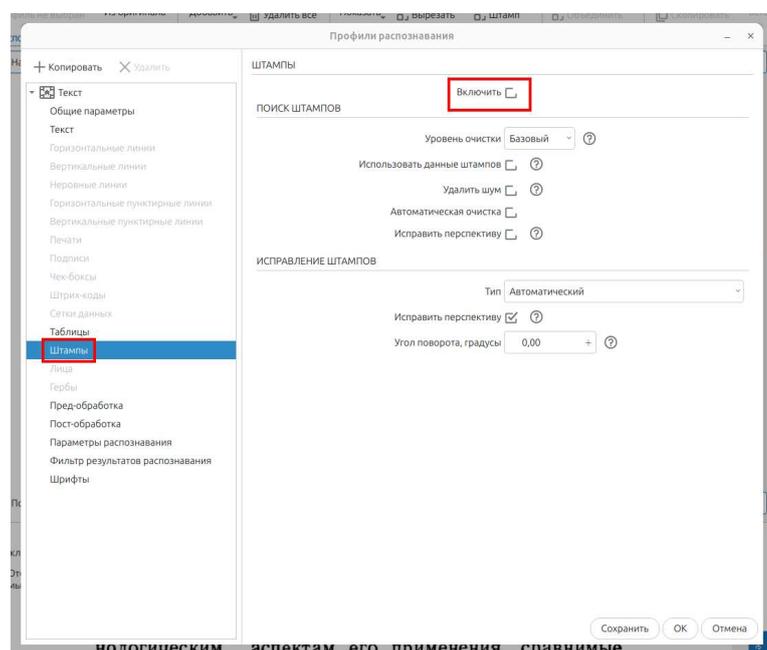
Автоповорот. Позволяет повернуть изображение, если система распознавания считает его ориентацию неверной.

Угол поворота, градусы. Угол в градусах, на который необходимо повернуть изображение перед обработкой.

Нормализация. Позволяет обрезать изображение относительно найденного лица (например, для паспорта).

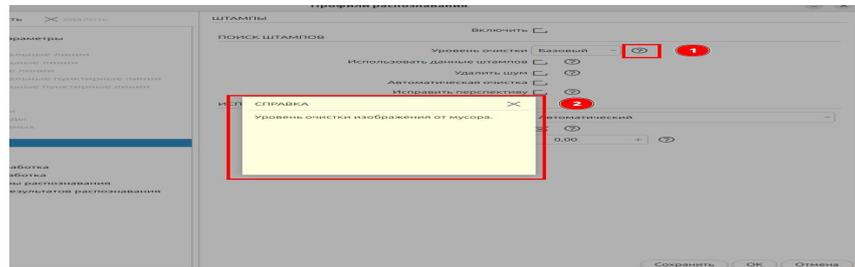
Исправить перспективу. Устранение на изображении перспективного схождения линий, которые в реальности должны быть параллельными.

Разделы, которые в списке подсвечены бледно серым цветом, означают, что они не активны и эти параметры не участвуют в извлечении данных. Для того, чтобы включить и настроить нужные объекты для поиска (текст, горизонтальные линии, вертикальные линии, неровные линии, горизонтальные пунктирные линии, вертикальные пунктирные линии, печати, подписи, чек-боксы, штрих-коды, сетки данных, таблицы, штампы, лица, гербы) необходимо перейти в нужную вкладку и выбрать Включить (Рис.43).



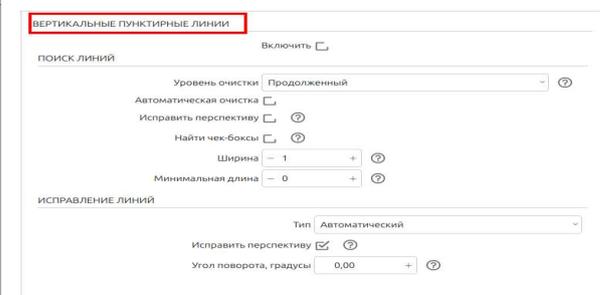
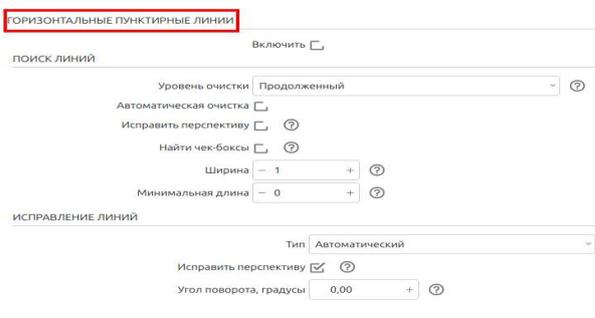
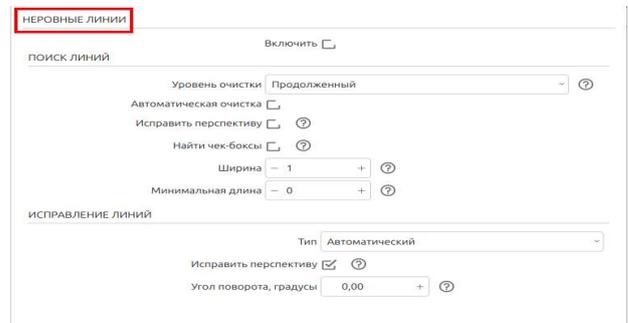
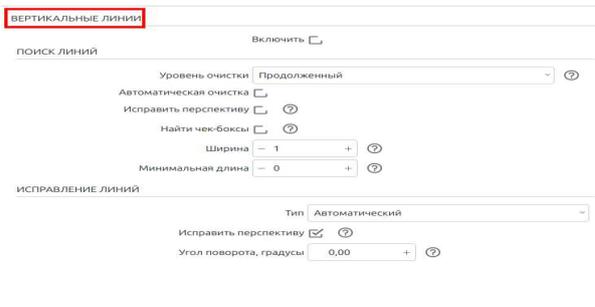
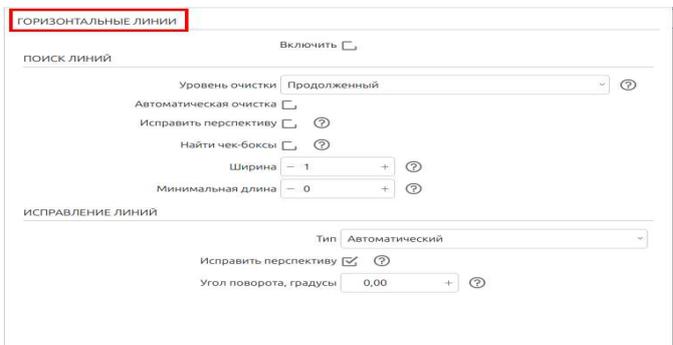
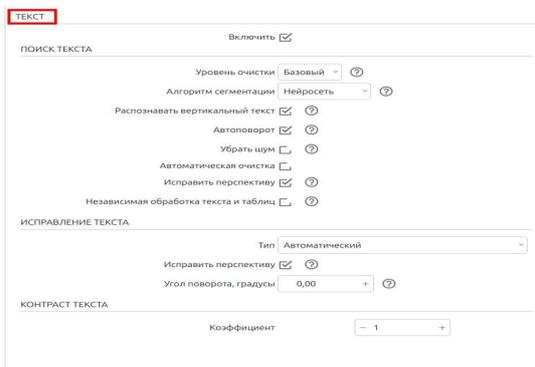
(Рис.58 Включение поиска нужного объекта)

Далее, с помощью подсказок, вызываемых по кнопке Справка (Рис.59) выбрать нужные параметры.



(Рис.59 Вызов справки по нужному параметру)

Разные разделы поиска объектов или групп объектов содержат разные настройки (Рис.60)



ПЕЧАТИ

Включить

ПОИСК ПЕЧАТЕЙ

Уровень очистки: Базовый

Удалить шум

Автоматическая очистка

Исправить перспективу

Найти треугольник в центре

Минимальный радиус: 0

Минимальный коэффициент: 0,00

Максимальный коэффициент: 0,30

ИСПРАВЛЕНИЕ ПЕЧАТЕЙ

Тип: Автоматический

Исправить перспективу

Угол поворота, градусы: 0,00

ПОДПИСИ

Включить

ПОИСК ПОДПИСЕЙ

Уровень очистки: Базовый

Удалить шум

Использовать данные штампов

Автоматическая очистка

Исправить перспективу

ИСПРАВЛЕНИЕ ПОДПИСЕЙ

Тип: Автоматический

Исправить перспективу

Угол поворота, градусы: 0,00

ЧЕК-БОКСЫ

Включить

ПОИСК ЧЕК-БОКСОВ

Уровень очистки: Обрамляющий прямоугольник текста

Удалить шум

Автоматическая очистка

Исправить перспективу

Доля от максимального размера изображения: 0,00

Использовать данные однострочных таблиц

ИСПРАВЛЕНИЕ ЧЕК-БОКСОВ

Тип: Автоматический

Исправить перспективу

Угол поворота, градусы: 0,00

ШТРИХ-КОДЫ

Включить

ПОИСК ШТРИХ-КОДОВ

Уровень очистки: Базовый

Удалить шум

Автоматическая очистка

Исправить перспективу

Повернуть

ИСПРАВЛЕНИЕ ШТРИХ-КОДОВ

Тип: Автоматический

Исправить перспективу

Угол поворота, градусы: 0,00

СЕТКИ ДАННЫХ

Включить

ПОИСК СЕТОК ДАННЫХ

Уровень очистки: Базовый

Удалить шум

Автоматическая очистка

Исправить перспективу

ИСПРАВЛЕНИЕ СЕТОК ДАННЫХ

Тип: Автоматический

Исправить перспективу

Угол поворота, градусы: 0,00

ТАБЛИЦЫ

Включить

ПОИСК ТАБЛИЦ

Уровень очистки: Базовый

Использовать пунктирные линии

Удалить шум

Автоматическая очистка

Исправить перспективу

ИСПРАВЛЕНИЕ ТАБЛИЦ

Тип: Автоматический

Исправить перспективу

Угол поворота, градусы: 0,00

ШТАМПЫ

Включить

ПОИСК ШТАМПОВ

Уровень очистки: Базовый

Использовать данные штампов

Удалить шум

Автоматическая очистка

Исправить перспективу

ИСПРАВЛЕНИЕ ШТАМПОВ

Тип: Автоматический

Исправить перспективу

Угол поворота, градусы: 0,00

ЛИЦА

Включить

ПОИСК ЛИЦ

Уровень очистки: Базовый

Удалить шум

Автоматическая очистка

Исправить перспективу

ИСПРАВЛЕНИЕ ЛИЦ

Тип: Автоматический

Исправить перспективу

Угол поворота, градусы: 0,00

ГЕРБЫ

Включить

ПОИСК ГЕРБОВ

Уровень очистки: Базовый

Размер картинки: 640

Порог достоверности: 0,25

Удалить шум

Автоматическая очистка

Исправить перспективу

ИСПРАВЛЕНИЕ ГЕРБОВ

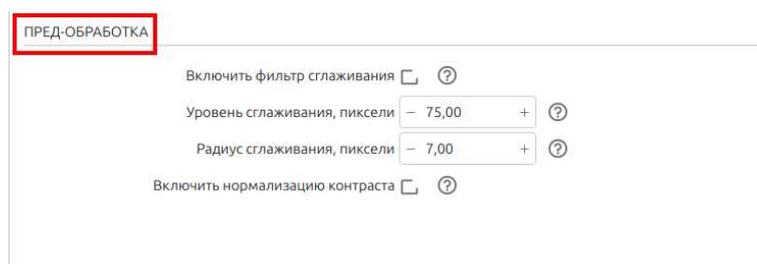
Тип: Автоматический

Исправить перспективу

Угол поворота, градусы: 0,00

(Рис. 60 Настройки разных разделов поиска объектов)

В профиле распознавания можно использовать пред-обработку с помощью настроек соответствующего раздела (Рис.61).



(Рис.61 Настройки пред-обработки)

В разделе Пред-обработка следующие настройки:

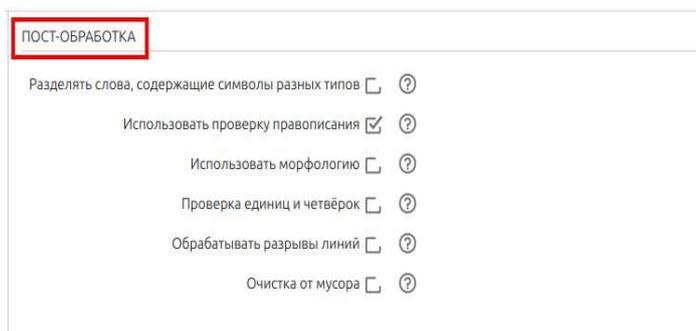
Включить фильтр сглаживания. Позволяет использовать фильтр при распознавании.

Уровень сглаживания, пиксели. Это интенсивность округления края каждого отдельного символа.

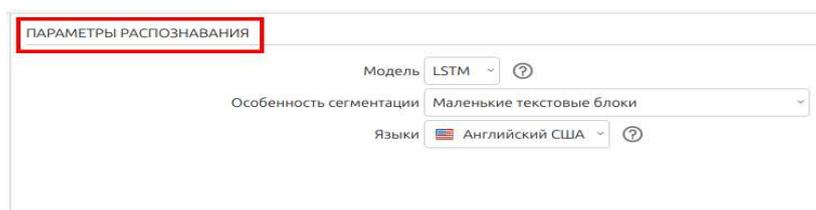
Радиус сглаживания, пиксели. Это величина радиуса округления края каждого отдельного символа.

Включить нормализацию контраста. Это обработка контрастного ореола около каждого символа.

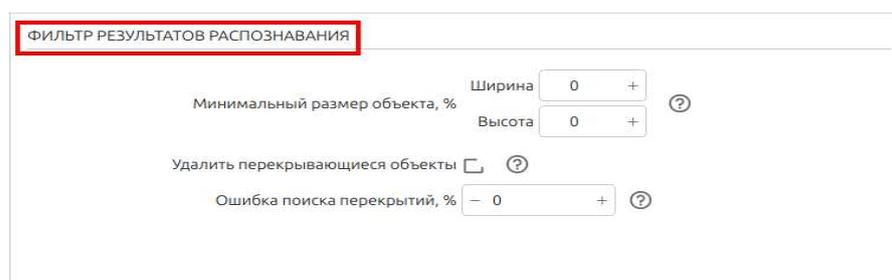
В настройки профиля распознавания можно включить пост-обработку (Рис.62), изменить стандартные Параметры распознавания (Рис.63), отфильтровать результаты распознавания (Рис.64) и указать параметры шрифта для распознанного текста (Рис.65).



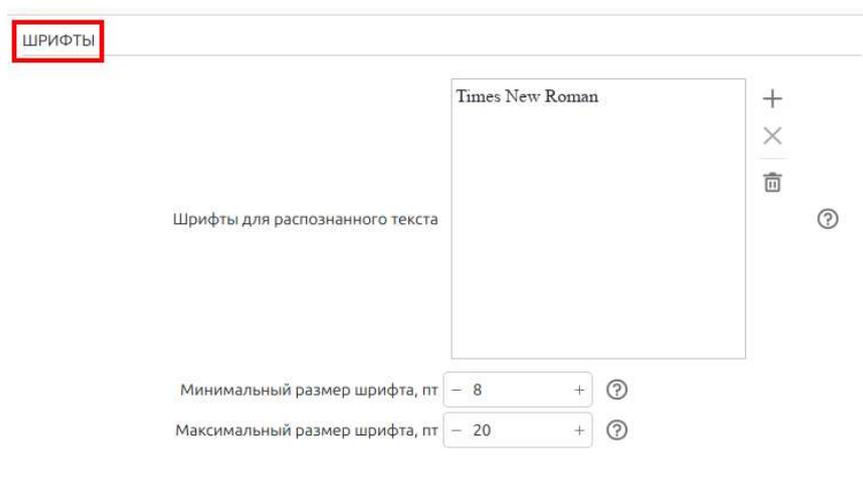
(Рис.62 Настройки пост-обработки)



(Рис. 63 Настройки параметров распознавания)



(Рис.64 Настройки фильтра результатов распознавания)



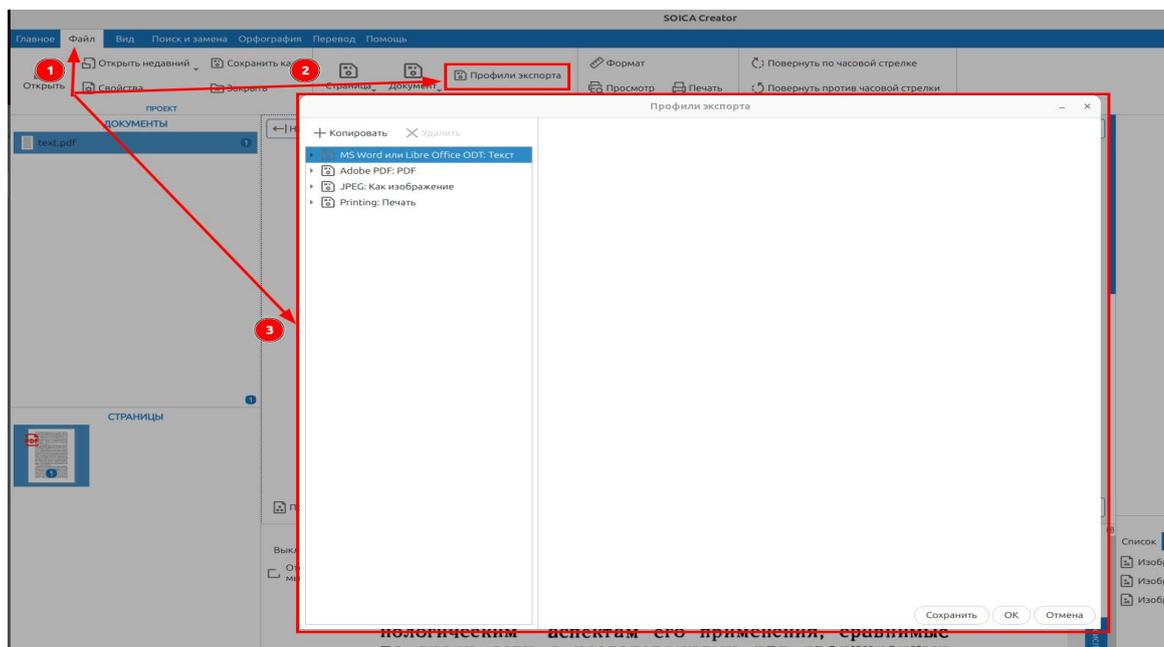
(Рис.65 Настройки шрифта для распознанного текста)

После выполнения настройки профиля распознавания его необходимо сохранить, нажав кнопку Сохранить.

5. Профили экспорта

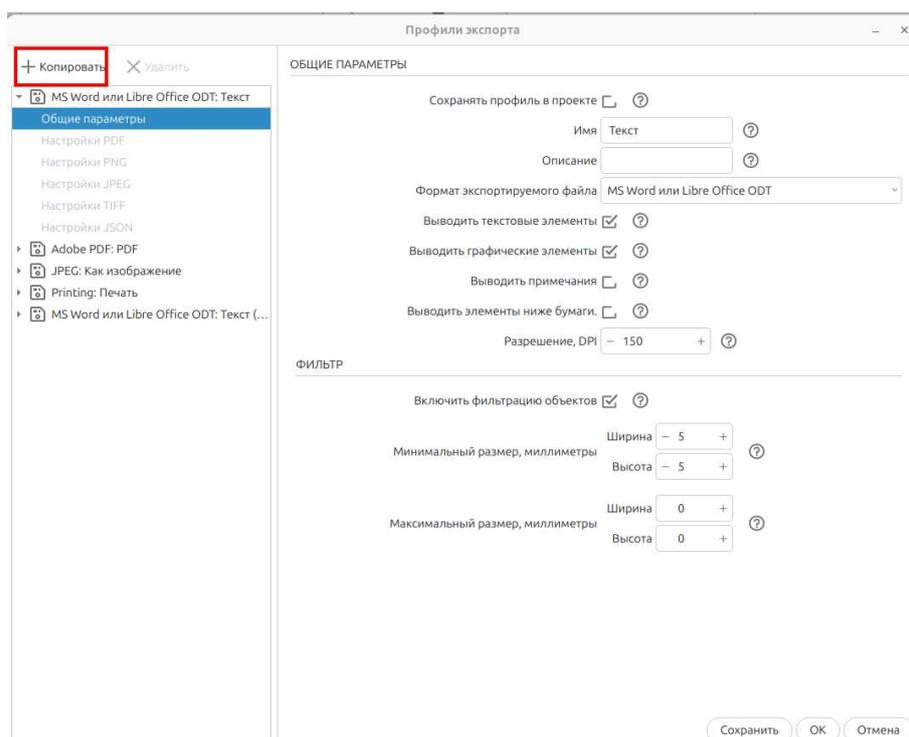
Экспортирование – процесс вывода найденных данных в заданном виде.

Чтобы создать, настроить или изменить профиль экспорта необходимо перейти в меню Файл, в разделе экспорт нажать на Профили экспорта. Произойдет переход к настройкам профилей экспорта (Рис.66).



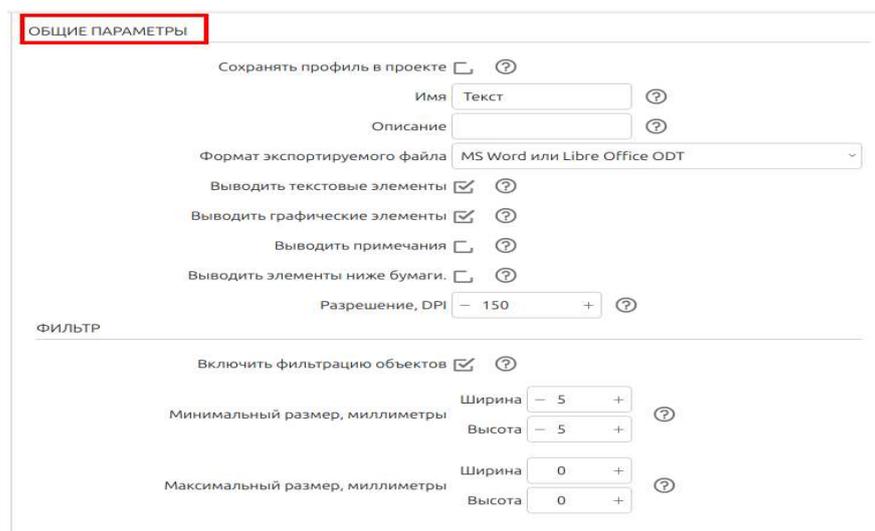
(Рис.66 Переход к настройкам профилей экспорта)

Далее можно скопировать один из профиль экспорта с такими же параметрами и отредактировать их (Рис.67).



(Рис.67 Общие настройки профилей экспорта)

В общих параметрах профиля экспорта (Рис.68) можно:



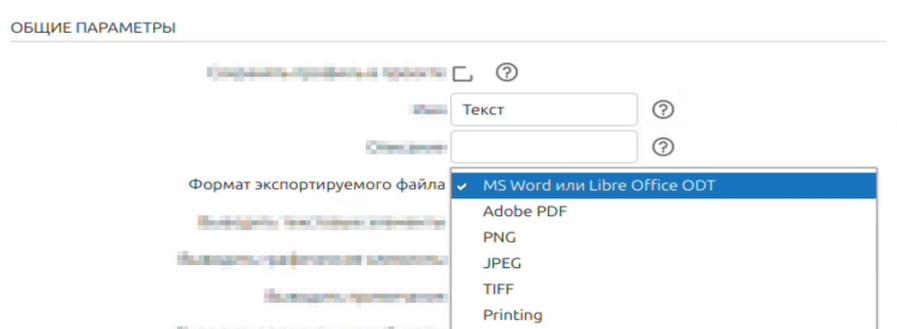
(Рис.68 Настройки общих параметров профиля экспорта)

Сохранить профиль в настройках приложения. Если этот параметр включен, профиль экспорта будет сохранён в настройках приложения. Иначе он будет сохранён в проекте.

Имя. Уникальное название профиля экспорта.

Описание. Комментарии к профилю экспорта.

Формат экспортируемого файла. Необходимо выбрать формат из выпадающего списка (Рис.69).



(Рис.69 Выбор формата экспортируемого файла)

Выводить текстовые элементы. Этот параметр позволяет сохранять в экспортируемом документе распознанные текстовые элементы.

Выводить графические элементы. Этот параметр позволяет сохранять в экспортируемом документе распознанные графические элементы.

Выводить примечания. Этот параметр позволяет сохранять в экспортируемом документе элементы, помеченные как примечания.

Разрешение, DPI. Значение этого параметра определяет визуальное качество документа и размер экспортированного файла. Чем больше это

значение, тем выше качество и больше размер файла. Рекомендуемый диапазон значений — от 150 до 600.

Включить фильтрацию объектов. Если этот параметр включен, в экспортируемом файле не будет объектов, для которых не выполняются условия фильтрации.

Минимальные размеры, миллиметры. Объекты размером меньше не войдут в экспортируемый документ. Нулевое значение означает отсутствие ограничения.



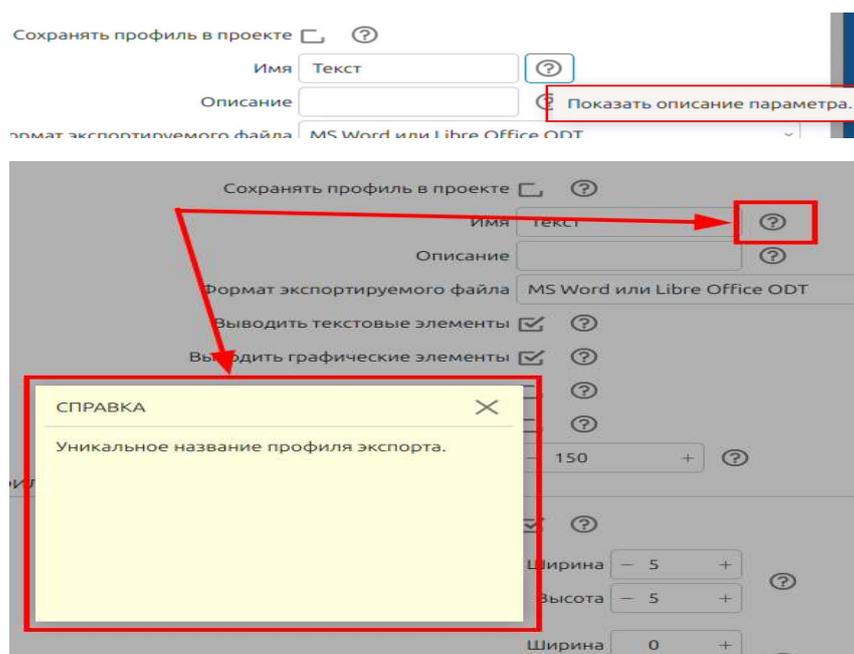
Максимальный размер, миллиметры. Объекты размером больше не войдут в экспортируемый документ. Нулевое значение означает отсутствие ограничения.



В нижней части настроек находятся кнопки Сохранить, Ок, Отменить.

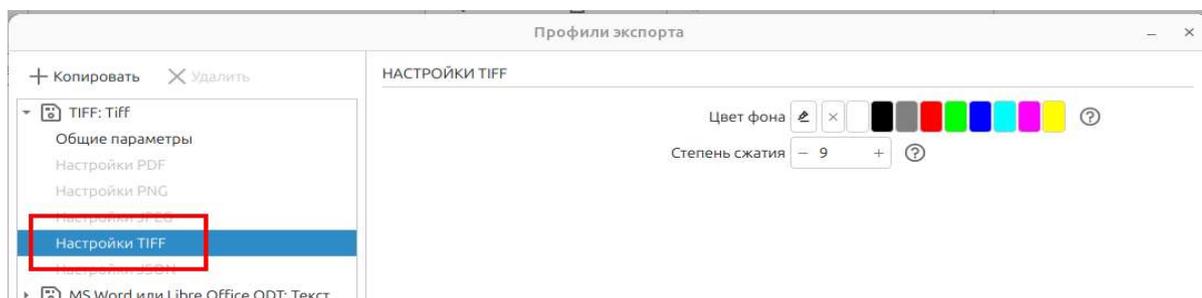
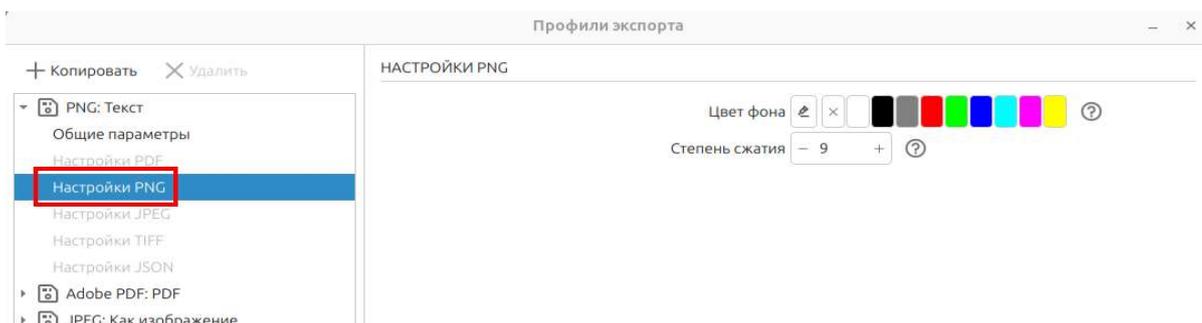


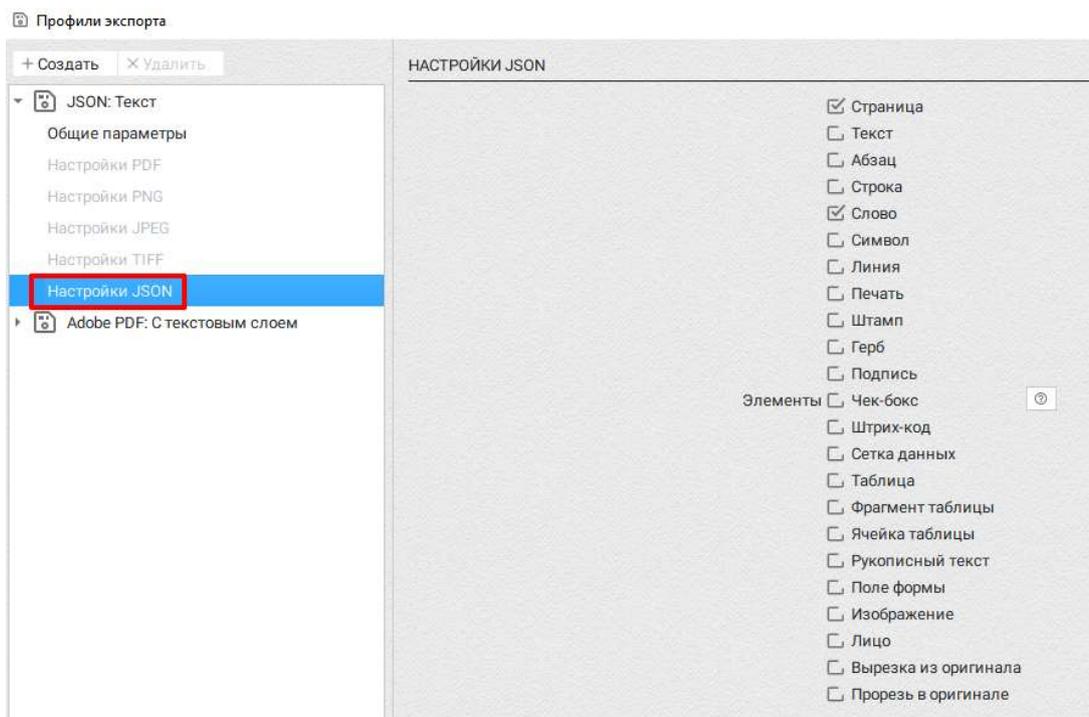
Кнопка  (Рис.70) позволяет выводить справку по выбранному параметру.



(Рис. 70 Кнопка вывода справки по параметру)

В зависимости от выбранного формата экспортируемого файла необходимо перейти в настройки данного формата (Рис.71), произвести настройки и сохранить профиль экспорта.





(Рис. 71 Настройки формата экспортируемого файла)

6. Системные требования программы.

Пользовательский компьютер, ноутбук (ubuntu 24.04 и другие дистрибутивы
Операционной системы Linux, 2 Гб RAM, 2 ядра CPU, 64-бита, 2Гб
свободного места на диске).

7. Установка и удаление программы.

7.1 Установка под дистрибутивы Linux.

Для установки пакета можно использовать утилиту **dpkg** или **rpm**, в зависимости от того, на чем базируется Linux-дистрибутив.

- 1) Скачайте или другим образом получите deb-пакет.
- 2) Перейдите в папку с установленным пакетом. Наш пакет находится в папке «Загрузки», чаще всего при скачивании программа попадает именно туда. Для входа в папку введите команду:

```
cd ~/Загрузки/
```

* Если ваш файл находится в другой папке, вместо /Загрузки/ введите нужную.

- 3) Установите пакет. Для этого введите:

- Для дистрибутивов на базе deb (AstraLinux) введите
`sudo dpkg -i soica_creator-7.0.0_amd64.deb`

Где `soica_creator-1.0.0_amd64.deb` — имя установочного файла.

- Для дистрибутивов на базе rpm (МСВСфера, РедОС) введите
`sudo rpm -i soica_creator-7.0.0_amd64.rpm`

4) Введите пароль для учетной записи.

5) Дождитесь установки программы.

8. Удаление программы.

8.1 Удаление в дистрибутивах Linux.

1. Открыть терминал
2. Ввести `sudo apt remove soica-creator`