



**SL SOFT**

×



**SOiCA**

# OCR-платформа SOiCA:

распознавание и обработка

текстовых и графических данных

# Знакомство с компанией



**SL Soft** (ООО «СЛ Софт») – разработчик российских бизнес-приложений для крупных государственных и корпоративных заказчиков.

Компания организована в апреле 2023 года для объединения усилий по развитию продуктов, уже существующих на рынке, а также созданию нового программного обеспечения.

Входит в ГК Softline.

## Преимущества

20+

лет опыта  
в создании  
продуктов

11

продуктовых  
линеек



централизованный  
сервис

## Акценты развития



Консолидация российской экспертизы в области разработки продуктов



Широкий продуктовый стек от одного вендора для покрытия всех потребностей заказчика



Развитие и расширение портфеля за счет продуктов с реальным опытом эксплуатации

# Продуктовая линейка



## Дополнительные сервисы SL Soft

- > ИТ-консалтинг
- > Доработка и кастомизация
- > Внедрение
- > Поддержка и сопровождение



Продукты SL Soft включены в Единый реестр российских программ для ЭВМ и БД Минцифры РФ.



# Примеры задач для OCR-платформ



## Кросс-отраслевые задачи

- Извлечение данных из ДУЛ
- Обработка входящих запросов
- Автоматизация ввода бухгалтерской и юридической документации
- Обработка кадровых документов
- Формирование реестров



## Банки и страхование

- Интеллектуальная обработка заявок на кредиты, лизинг, страховки, факторинг
- Контроль комплектности и качества
- Внешние и внутренние проверки (справочники 1С, SAP, налоговая, базы МВД, ограничения на выезд и др.)



## Государственные учреждения и муниципалитет

- Обработка обращений граждан
- Оцифровка ветхого фонда
- Единые реестры и базы данных
- Создание интеллектуальных хранилищ



## Образование и НИИ

- Архив учебных и научных материалов
- Обработка персональных документов



## Промышленность

- Хранилище проектно-сметной и технологической документации
- Обработка инструкций и сертификатов
- Контроль маркировки



## Энергетика

- Обработка заявок на технологическое присоединение
- Обработка бумажных протоколов испытаний

# Типовые обрабатываемые документы



## Финансы и бухгалтерия

- Торг12
- Счета
- Счета-фактуры
- Акты



## Юридический департамент

- Договоры и приложения
- Протоколы согласований
- Уставные документы
- Лицензии и патенты



## HR-департамент

- ДУЛ (паспорта, права)
- Кадровые документы
- Резюме
- Приказы и заявления



## Маркетинг и продажи

- Опросы покупателей
- Анкеты мерчандайзеров
- Ценники и чеки
- Заказы и заявки



## Делопроизводство

- Входящая корреспонденция
- Архивы и реестры
- Журналы регистраций



## Производственный отдел

- Проектная документация
- Чертежи
- Инструкции
- Сертификаты



**SOICA – единое решение для работы с любыми типами документов.**

# Почему SOiCA нужна даже при переходе на ЭДО

При получении документов неформализованного типа документов (договоры, приложения, спецификации, акты оказанных услуг и т.п.) **карточка документа содержит неполные данные**, поэтому ее приходится заполнять вручную.

Особую сложность представляют случаи, когда оригинал документа прислан в PDF без выделенного текстового слоя. **SOiCA может автоматически вычленять нужные данные и передавать их в карточку документа.**



Входящий Договор [redacted] Маркетолог ООО Авангард.pdf

Показать историю (3)

- Еремеева В. передал документ на подпись сотруднику: Павлов А. В.  
Сотрудник · 13 мар. 17:09
- Павлов А. В. подписал и отправил извещение о получении  
Генеральный Директор · 14 мар. 11:02
- Павлов А. В. подписал документ и завершил документооборот  
Генеральный Директор · 14 мар. 11:02

Написать контрагенту | Запросить аннулирование | Согласование | Скачать | Распечатать | Удалить

Подписан

Договор № 17742  
на оказание услуг в серви

г. Москва

Публичное акционерное общество [redacted]  
[redacted] в лице Руководителя группы выпуска региональных  
Шабера, Алексей Петрович, действующей на основании

Пример договора, где нет данных о номере, дате, сумме, а исходный файл – в PDF.

# Как работает SOICA



## 1 этап

### Поступление документов

- Скан-копии
- Фотографии
- Электронные документы

Любой формат		e-mail, папки, API
Любые каналы		jpeg, pdf, docx и др.

## 2 этап

### Улучшение качества изображения

- Исправление шумов
- Выравнивание
- Яркость/контраст
- Бинаризация и т.д.

18 фильтров

## 3 этап

### Распознавание данных

- Поиск слов и графических элементов на изображении
- Посимвольное распознавание

Собственный OCR-движок

## 4 этап

### Классификация документа

Определение типа документа. Разделение пакета документов на отдельные файлы

- Без шаблонов
- Без штрихкодов
- Без QR-кодов

## 5 этап

### Формирование карточки документа

- Работа логических цепочек по поиску и определению нужных атрибутов
- Заполнение карточки документа

## 6 этап

### Проверка комплектности

Автоматический запрос недостающих документов. Настраиваемый процесс работы с некомплектами.

## 7 этап

### Форматирование и валидация

- Проверки внутри комплектов
- Сверка данных с внешними источниками
- Математические, логические, форматные проверки
- Валидация оператором

## 8 этап

### Экспорт данных

- Дополнительные обработки: факсимиле, синтез, анонимизация
- Экспорт данных в любом формате (xml, json)

# Ключевые проблемы при работе с OCR-системами

-  Долго и сложно настраивать обработку новых типов документов.
-  Сложно настроить логику проверок внутри документа или комплекта документов, требуется работа программиста и написание кода.
-  Много данных приходится доносить вручную:
  - не распознается рукописный текст,
  - плохое качество распознавания данных из-за OCR-движка.

## В SOICA эти проблемы решены:

-  Используется новый OCR-движок.
-  Есть распознавание рукописного текста.
-  Визуальная (No-Code) среда настройки распознавания новых типов документов.
-  Настройка разных видов проверок в документах и комплектах документов без программирования.

# Новый OCR-движок SOICA

Движок SOICA точно детектирует нужные области. Это сокращает число ошибок распознавания – реже требуется проверка данных человеком. Больше автоматизации – меньше ручного ввода.



Новая версия



Старая версия



Лучше детектирование



Меньше шумов



Меньше ошибок распознавания

# Распознавание рукописного текста

Оригинал SOICA\_passport SOICA\_passport Рукописный

Фамилия: Еремеев

Имя: Сергей

Отчество: Анатольевич

Профессия: инженер

Образование: высшее

Текст: высшее профессиональное  
Доверие: 97.33656  
Инфо:

ВКЛАДЫШ В ТРУДОВУЮ КНИЖКУ  
(вкладыш без трудовой книжки  
недействителен)  
ВТ № 8435369

Фамилия: Ершищев

Имя: Сергей

Отчество: Анатольевич

Дата рождения: 05 сентября 1985

Образование: высшее профессиональное

Профессия, специальность: инженер

Дата заполнения: 08 апреля 2011 г.

Подпись владельца книжки: [подпись]

Подпись лица, ответственного за ведение трудовых книжек: [подпись]

## 1 Точность\*

- по словам: 80%
- по символам: 95%

## 2 Скорость обработки (1 страница):

- от 7 сек CPU
- 0,3 сек GPU

## 3 Поддерживается два языка:

- русский
- английский

\*Замерено на датасете

<https://www.kaggle.com/datasets/constantinwerner/cyrillic-handwriting-dataset/data>

# Легкая настройка распознавания НОВЫХ ТИПОВ ДОКУМЕНТОВ

## No-Code среда настройки:

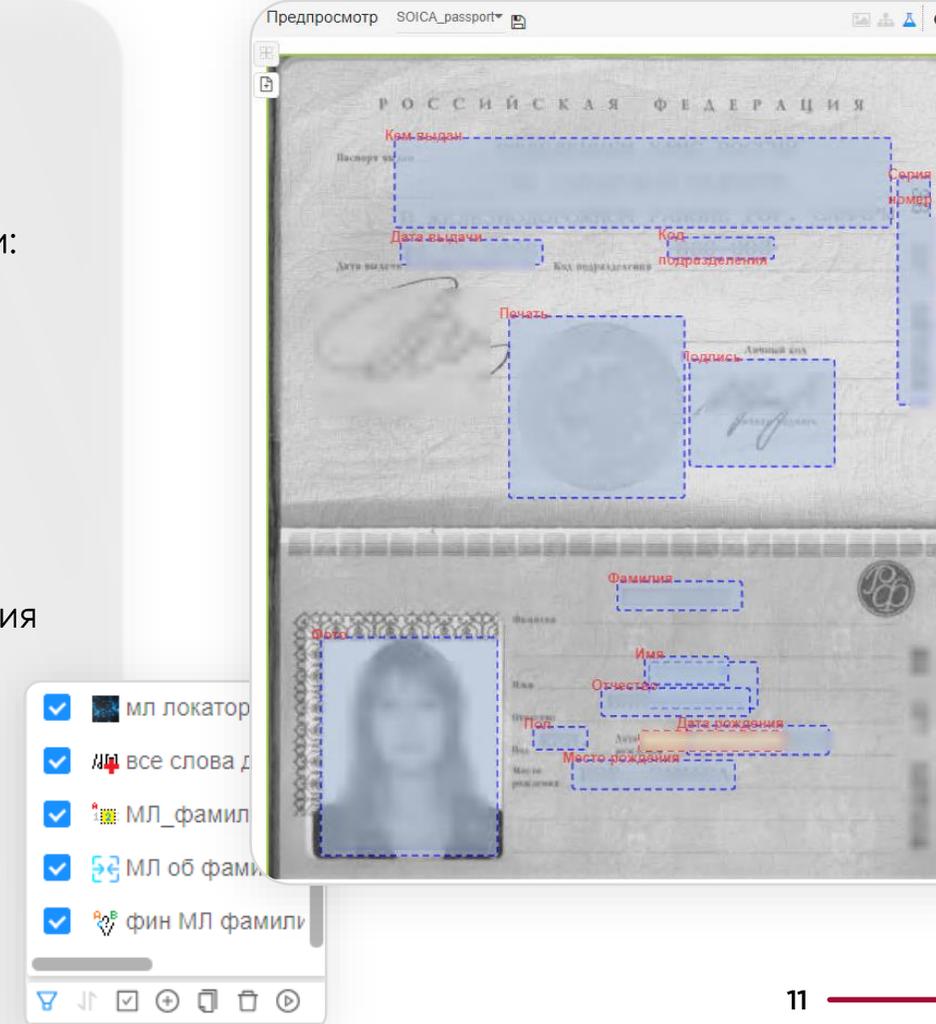
- визуальный интерфейс,
- настройка обработки любых документов без программирования.

## Это позволяет:

- снизить стоимость владения системой (ниже требования к персоналу -> легче найти, дешевле),
- сократить время на внедрение,
- снизить зависимость от разработчика и специалистов поддержки.

## ML- локатор (новый инструмент среды):

- простая настройка обработки документов с помощью нейросети: целевые данные размечаются на примерах прямо в среде настройки SOICA,
- еще проще настройка распознавания новых типов документов,
- еще больше снижаются требования к персоналу,
- повышается качество детектирования данных.



# Настройка проверок в документах и комплектах документов без программирования



## Математические проверки цифр и сумм

Проверка на правильность «Всего» в соответствии со стоимостью, количеством, суммой НДС

## Сверка данных с внешними базами

ЕГРЮЛ, база паспортов МВД и т.п.

## Сверка внутри комплекта

Например, совпадает ли ФИО контрагента в договоре, паспорте, правах и т.д.

Пример проверки суммы, где выявлено несоответствие по формуле «Сумма без НДС + Сумма НДС = Сумма с НДС»

Всего к оплате с налогом: 69800.00

Всего к оплате сумма налога (НДС): 111633.33

Всего к оплате стоимость без НДС: 58166.67

Текст: 111633.33  
Доверие: 0  
Инфо: проверьте суммы

единицу измерения	(работ, услуг), имущественных прав без налога - всего	число сумма акциза	вая ставка	предъявляемая покупателю	услуг), имущественных прав с налогом - всего	ци
4	5	6	7	8	9	
58 166,67	58 166,67	без акциза	20%	11 633,33	69 800,00	
	58 166,67	X		11 633,33	169 800,00	

# SOICA: резюме

## Решаемые задачи

- Оцифровка входящего потока
- Оцифровка ветхого фонда
- Оцифровка бумажных документов
- Атрибутирование электронных документов
- Сравнение документов
- Модификация изображений

## Ключевые преимущества

**95%**

распознавания  
без участия человека

**Быстрое**  
внедрение и **простая**  
настройка

Низкая стоимость  
обработки документа:  
**от 1 до 3 руб.**  
за страницу



**Выгодное**  
владение

- **Бессрочная** лицензия
- **Безлимитное** количество пользователей
- **Неограниченное** количество точек сканирования
- **Бесплатная** настроенная среда

**NEW**

Распознавание  
**рукописного текста**

**Удобное управление**  
и модификация  
шаблонов

**Low-Code/No-Code**  
инструменты. Настройка  
сценариев без кода

# SOiCA и партнерские решения



Синергия продуктов SOiCA и решений **технологических** партнеров дают заказчикам **дополнительные преимущества** в области автоматизации документооборота и сторонних бизнес-процессов, значительно ускоряя их и сокращая нагрузку на персонал



# Пример синергии продуктов SL Soft



# Проектный опыт + кейсы



Финансы

ПРОЛЕУМ

NDA

Банк, ТОП-5



Транспорт, логистика

САЛАИР



ЕвроТрейдинг



IT, Телеком

Софтлайн  
Аутсорсинг  
by Sofiline

NDA

Провайдер цифровых  
услуг и сервисов, ТОП-5



Образование

Sk Сколково



Ритейл

СИТИЛИНК  
все лучшее для тебя



Энергетика

NDA

Энергетическая  
компания РФ



**SL SOFT**

×



**SOICA**

**Спасибо  
за внимание**



**Команда SOICA**

[info@slsoft.ru](mailto:info@slsoft.ru)