

**Как работать с неотредактируемыми документами: решения Content AI для личного и корпоративного использования**

# Спикеры



Ирина Семшова, руководитель отдела по работе с партнерами



Андрей Малахов, ведущий менеджер по продажам



Екатерина Ножнёва, руководитель проектов

# О компании

Content AI – российский разработчик ИИ-решений в области интеллектуальной обработки и поиска информации.

Компания использует всемирно признанные технологии распознавания текста, классификации документов и обработки естественного языка.

Решения Content AI помогают заказчикам из различных сфер и государственным организациям автоматизировать бизнес-процессы работы с документами.

**30 лет**

*технологического опыта  
на российском рынке*

**5 продуктов**

*имеют официальную ИИ-маркировку в  
реестре отечественного ПО*

**140+**

*сотрудников в компании, 65%  
инженеров и разработчиков*

**МИНЦИФРЫ**

*Content AI имеет  
аккредитацию Минцифры РФ*

**РЕЕСТР**

*Все продукты входят в  
реестр отечественного ПО  
Минцифры РФ*

**200+**

*партнеров  
на территории России и  
СНГ*

# Ключевые продукты и решения

## Массовое ПО



ContentReader PDF



Lingvo by Content AI



## ПО для разработчиков



ContentReader Engine



PassportReader SDK



## Корпоративное ПО



ContentCapture



ContentReader Server



Content AI Intelligent Search



# ContentReader PDF 15

Многофункциональный редактор для решения любых задач с PDF и бумажными документами

ContentReader PDF упрощает оцифровку, поиск, редактирование документов, обмен файлами и совместную работу над любыми типами документов, позволяя работать с PDF так же просто и удобно, как с привычными всем текстовыми документами



# Ключевые возможности ContentReader PDF

## 1. Работа с PDF

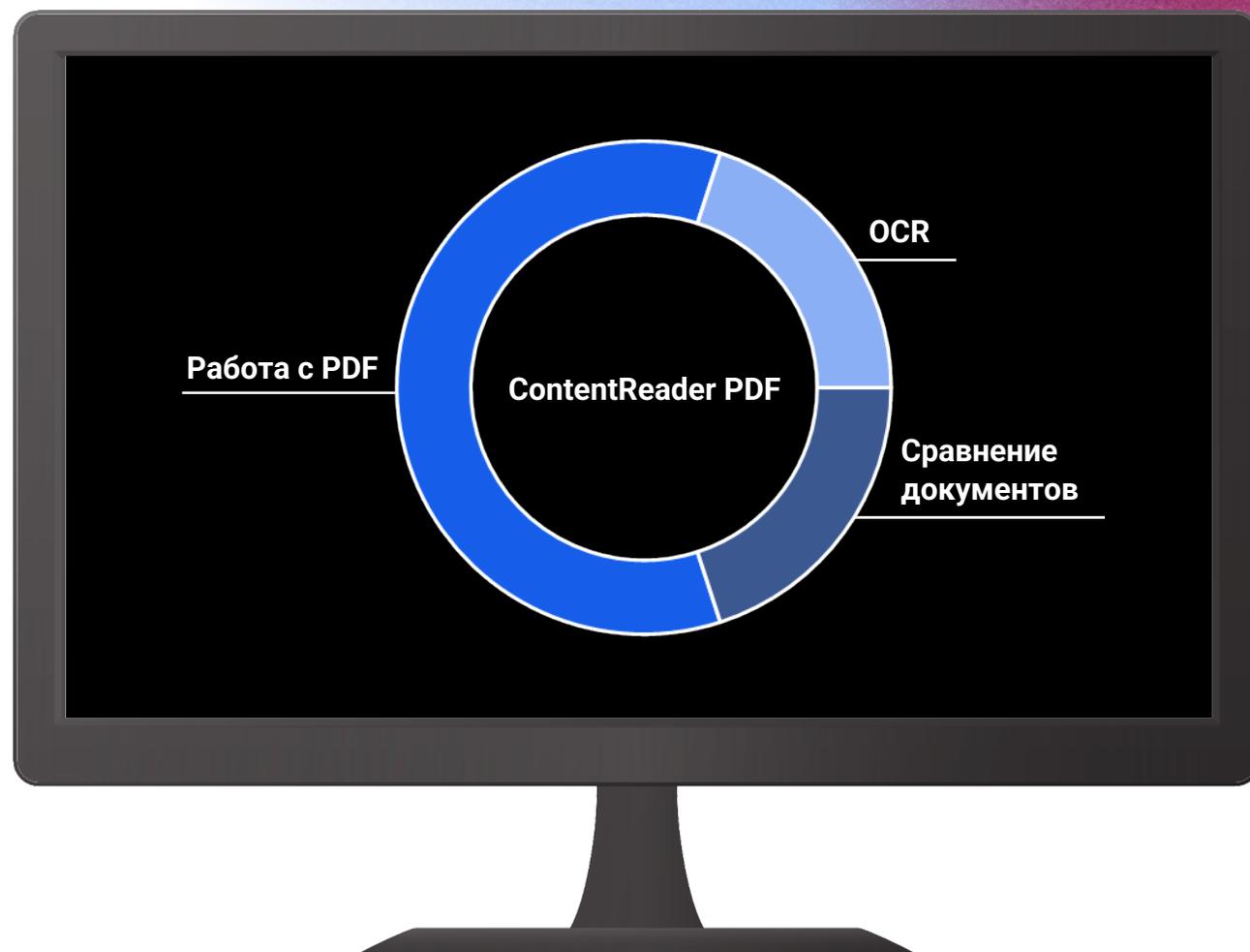
- Просмотр
- Редактирование
- Создание
- Комментирование
- Объединение
- Защита

## 2. OCR

- Сканирование
- Конвертирование
- Поиск
- Копирование текста

## 3. Сравнение документов

- Просмотр различий



# Сценарии использования для каждого сотрудника



## Секретари, офис-менеджеры

Обработка входящей корреспонденции, редактирование PDF-документов, создание шаблонов документов



## Менеджеры по продажам

Редактирование бумажных и PDF-документов, создание PDF из файлов разных форматов, сравнение документов



## Маркетологи и аналитики

Создание PDF из файлов различных форматов, конвертация в цифровые форматы, согласование и подготовка макетов



## Бухгалтеры

Извлечение информации, уменьшение размеров PDF-файлов, операции со страницами, удаление информации



## Юристы

Сравнение двух версий документа, удаление конфиденциальной информации, защита и подпись документов



## Научные сотрудники

Создание электронных копий научных статей, учебных материалов, подготовка отчетов на основе информации из разных источников

# Технические характеристики

## ПОДДЕРЖИВАЕМЫЕ ФОРМАТЫ

### Входные форматы файлов:

- PDF, включая PDF/A
- Графические форматы: TIFF, JPEG, PNG, BMP, PCX, GIF
- Текстовые форматы: DOCX, XLSX, PPTX, HTML, RTF, TXT, ODT, ODS, ODP

### Форматы экспорта:

- Графические форматы: TIFF, JPEG, JPEG 2000, JBIG2, PNG, BMP, PCX
- Текстовые форматы: DOCX, XLSX, PPTX, HTML, RTF, TXT, CSV, ODT
- Электронные книги: EPUB, FB2

## ПОДДЕРЖИВАЕМЫЕ ОС

**ContentReader PDF для Linux работает под управлением российских ОС и совместим с:**

- Astra Linux Special Edition 1.7, Common Edition 2.12.46
- Astra Linux 1.8
- РЕД ОС 7.3, 8
- Альт Рабочая станция 10, Рабочая станция К 10
- Атлант Сборка Alcyone версия 2022.02
- AlterOS 7.5
- МСВСфера 9.2
- ОС «ОСнова» с обновлениями 2.7
- Uncom OS



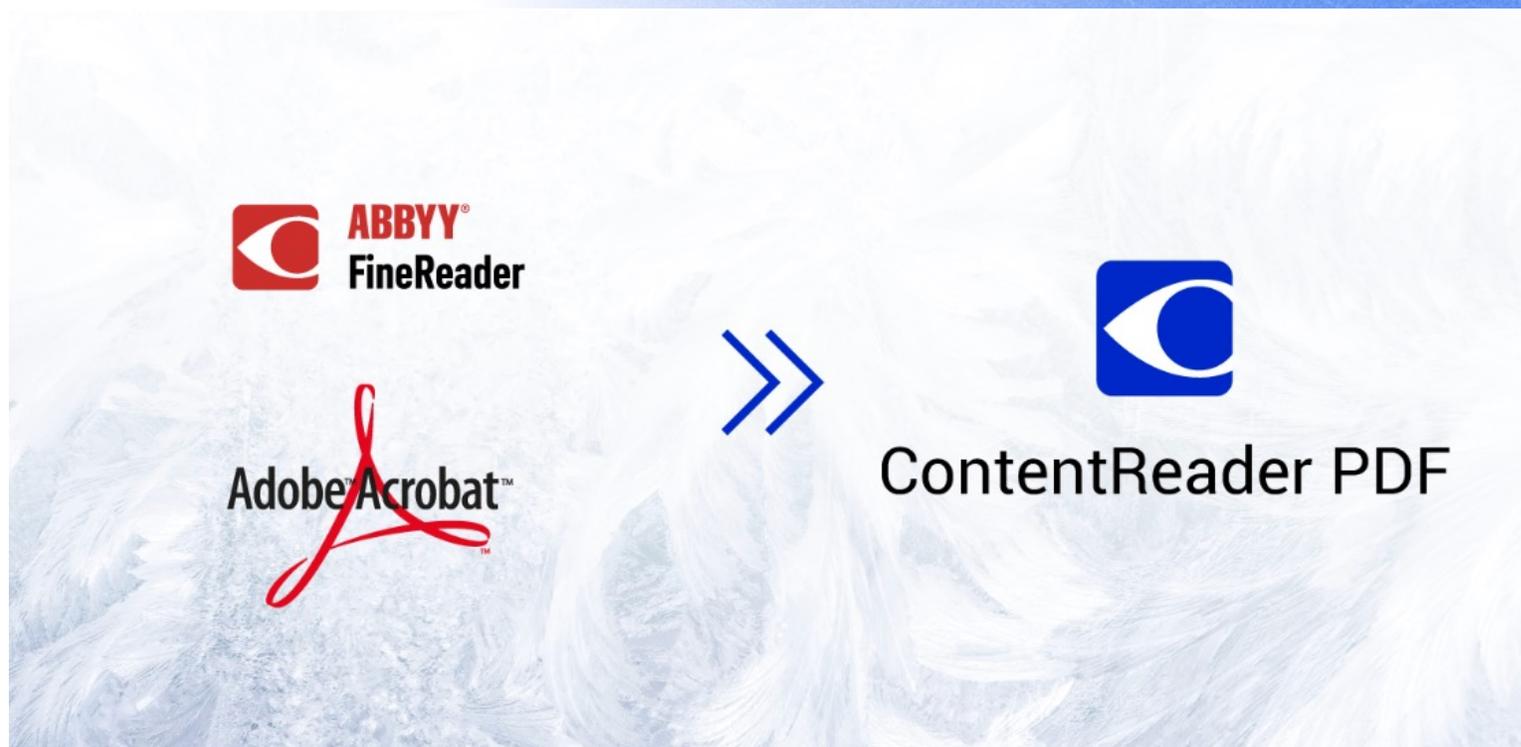
# 7 преимуществ решения по сравнению с конкурентами

- 1 Наиболее полная функциональность для работы с PDF-документами на Linux
- 2 Высокая точность и скорость работы благодаря лучшим AI-технологиям на рынке
- 3 Возможность легко копировать таблицы из PDF с сохранением их структуры
- 4 Простой и понятный интерфейс, созданный на основе многолетнего опыта и протестированный нашими UX-специалистами
- 5 Редактирование текста на уровне абзаца, а не на уровне отдельных слов. Это позволяет работать с PDF, как с обычным текстовым документом
- 6 Быстрая и точная конвертация отсканированных документов в редактируемые форматы
- 7 Техническая поддержка, подробная справка, возможность доработки решения под запросы клиента

# Импортозамещение

Полный инструментарий для работы с PDF, соответствующий требованиям цифровой среды

Решить разные задачи можно в одной программе. Высокое качество распознавания и интуитивно понятный интерфейс упрощают работу с любыми документами и текстами



**Цель:** перевести 5000 сотрудников ПАО «Ростелеком» по всей России на отечественный редактор PDF

# Импортозамещение

## Задачи

- Оперативно заместить зарубежное ПО (ABBYY FineReader PDF, Adobe Acrobat Pro, Foxit PDF Reader)
- Сохранить функциональность и высокое качество, иметь возможность масштабного развертывания ПО

## Результаты

**Более 5 тысяч**

сотрудников ПАО «Ростелеком» по всей России используют единый редактор — ContentReader PDF

## Бесшовная миграция

Переход осуществлен за 4 месяца без потерь в наборе функций и качества. ИТ-специалистам больше не нужно тратить время на обновление нескольких ПО

## Масштабирование

Пул сетевых лицензий Concurrent расширен в 3 раза. Централизованный сервер лицензирования позволяет контролировать лицензионный портфель из одной точки



**Цель:** осуществить полноценную замену импортного ПО для распознавания и обработки документов на российский аналог

# Импортозамещение

## Задачи

- Соответствовать всем функциональным требованиям пользователей к продукту
- Гарантировать надежность и качество работы софта при масштабном развертывании продукта

## Результаты

**Более 40  
предприятий**

Группы ПАО «Транснефть»  
используют ContentReader PDF

## Совместимость

Редактор прошел все этапы тестирования экспертами компании и продемонстрировал соответствие функциональным и техническим требованиям

## Полноценная замена

В процессе тестирования решено было заменить сразу два иностранных продукта на многофункциональный редактор ContentReader PDF



**Транснефть**



# Отзывы

*«С PDF-документами работают **специалисты многих подразделений АЛРОСА**. Для эффективного обеспечения бизнес-процессов и документооборота нам необходим **надежный и безопасный инструмент**, который позволит решать разнообразные задачи с файлами этого формата. Мы отметили **универсальность и многофункциональность редактора ContentReader PDF**. Продукт не уступает **иностранному софту**, который использовался ранее, и позволяет **закрывать все потребности профильных специалистов**»*

**Софья Таблина**

Руководитель проекта отдела  
по импортозамещению АЛРОСА

***ContentReader Server***  
***OCR сделать проще простого***  
***Или нет...?***

# Распознавание

**OCR** – Оптическое распознавание текста  
(200+ языков)

- European (Latin, Cyrillic, Armenian, Greek alphabets)
- CJK: Chinese (Simplified и Traditional), Japanese, and Korean
- Thai, Vietnamese, Hebrew, Burmese, Bangla
- Arabic

da

ထိုအခါဘုရားရှင်ဇော်တော်  
တို့သည်အလှူငွေဖြင့်လေ  
ကမ်းယံလည်းထိုအခါပျက်  
ပြိုကမ်းယံငွေလှူ၏။ကံ  
ရန်လေးဘုရားကိုအလှူ  
လေ၏။နမူနာပေးသော  
နှင်လှူလေ၏။မောင်လည်း  
လှူ၏။ထိုအခါပျက်ဘုရား  
တို့သည်တရားမင်းလှူ  
သောကံကံကံကံကံကံကံ

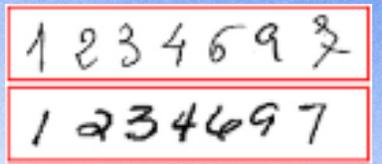
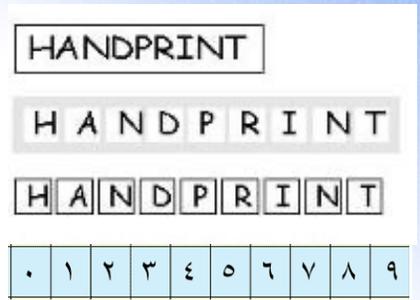
**BCR** – Распознавание баркодов



**OMR** – Распознавание чекмарок



**ICR** – Распознавание рукопечатных символов  
(130+ языков)



**HTR** – Распознавание рукописного текста



# Полнотекстовое распознавание



ContentReader PDF



ContentReader Server



ContentReader Engine



Переиспользование текста  
ContentReader (PDF, Server, Engine)



Конвертация формата  
ContentReader (PDF, Server, Engine)



Сравнение версий  
ContentReader (PDF, Engine)



Индексация/поиск  
IntelligentSearch

## ЛИЦЕНЗИОННЫЙ ДОГОВОР № 111111/23Д на право использования программы для ЭВМ «Контур.Диалок»



16.01.2023

Екатеринбург  
Акционерное общество «Производственная фирма «СКБ Контур», именуемое в дальнейшем Лицензиар, в лице генерального директора Сроных Михаила Юрьевича, действующего на основании Устава, с одной стороны, и АО «КОМБИНАТ», именуемое в дальнейшем Лицензиат, в лице Исполнительного Директора Ивановой Елены Ивановны, действующей на основании Устава, с другой стороны, именуемые в дальнейшем также Стороны, заключили настоящий Договор о нижеследующем.

### 1. Термины и определения

1.1. Контур.Диалок (иное наименование – «Диалок») – результат интеллектуальной деятельности – программа для ЭВМ «Контур.Диалок» (в том числе интеграционные модули и иные модули, предусмотренные прайс-листом, позволяющие Лицензиату использовать дополнительный функционал Контур.Диалока), размещенная на сервере Лицензиара и предназначенная для обеспечения юридически значимого электронного документооборота между хозяйствующими субъектами.

1.2. Спецификация (Приложение № 1 к Лицензионному договору, если в период действия Лицензионного договора Стороны согласовывают финансовые условия путем подписания Спецификации) – документ, содержащий информацию о стоимости и комплекте предоставляемых Лицензиату неисключительных прав использования программ для ЭВМ и оказываемых услуг/выполняемых работ. Одновременно с подписанием Лицензионного договора Стороны подписывают Спецификацию № 1. При продлении Лицензионного договора и/или при необходимости предоставления Лицензиату дополнительных неисключительных прав на использование программ для ЭВМ и/или услуг/работ, Стороны оформляют дополнительные Спецификации (также являющиеся Приложением № 1 к Договору), которым последовательно присваиваются порядковые номера. Количество Спецификаций к Лицензионному договору не ограничено.

1.3. Сублицензионный договор (Приложение к Лицензионному договору) – договор, устанавливающий порядок передачи и использования СКЗИ. Является офертой, не требующей подписания Сторонами, полный и безоговорочный акцепт которой Лицензиатом является существенным условием Лицензионного договора.

1.4. Квалифицированный сертификат (далее – Сертификат) – документ, выданный аккредитованному в соответствии с Федеральным законом Российской Федерации «Об электронной подписи» удостоверяющим центром и подтверждающий принадлежность ключа проверки электронной подписи владельцу сертификата ключа проверки электронной подписи.

1.5. Конечный пользователь – юридическое лицо – аффилированное лицо Лицензиата, в интересах которого действует Лицензиат в рамках Лицензионного договора. Ответственность за ознакомление Конечного пользователя с текстом Лицензионного договора лежит на Лицензиате. Конечный пользователь идентифицируется уникальной комбинацией ИНН-КПП. Все условия Лицензионного договора, за исключением условий, установленных пп. 5.3.1, а также разделами 7-8, 11 распространяются в равной степени как на Лицензиата, так и на Конечных пользователей. Конечные пользователи обладают теми же правами и несут те же обязанности по Лицензионному договору, что и Лицензиат. Во всех Лицензионном договоре, за исключением пп. 5.3.1, а также разделами 7-8, 11, при упоминании Лицензиата подразумевается также Конечный пользователь.

1.6. Список конечных пользователей (Приложение № 2 к Лицензионному договору) – документ, содержащий информацию о Конечных пользователях, являющийся неотъемлемой частью Лицензионного договора в том случае, если Лицензиат в рамках Лицензионного договора действует не только в своих интересах, но и в интересах Конечных пользователей. В случае необходимости приобретения Лицензиатом или Конечными пользователями лицензий на использование модулей Контур.Диалока и/или лицензий на иное программное обеспечение, а также оказания разовых услуг или выполнения работ, предусмотренных прайс-листом, такие лицензии, услуги и работы приобретаются отдельно в отношении Лицензиата и/или каждого Конечного пользователя.

1.7. Доверенный оператор электронного документооборота – функции Лицензиара, установленные приказом Министерства финансов Российской Федерации «Об утверждении Порядка выставления и получения счетов-фактур в электронной форме по телекоммуникационным каналам связи с применением усиленной квалифицированной электронной подписи».

1.8. Правила по обеспечению информационной безопасности на рабочем месте – документ, составленный Лицензиаром на основании положений действующего законодательства Российской Федерации в области применения и использования электронной подписи и СКЗИ и обязательный для ознакомления всеми специалистами Лицензиата, работающими с использованием СКЗИ. Актуальная редакция правил публикуется на сайте <https://ca.skbkontr.ru>. Подписанием Лицензионного договора Лицензиат подтверждает выполнение Лицензиаром обязанности, установленной частью 4 статьи 18 Федерального закона «Об электронной подписи».

1.9. Тарифный план – совокупность предоставляемых Лицензиаром неисключительных прав использования программ для ЭВМ и оказываемых услуг/выполняемых работ. Состав тарифного плана определяется прайс-листом.

1.10. Прайс-лист – документ (неотъемлемая часть Лицензионного договора), отражающий ценовую политику Лицензиара и содержащий сведения о тарифных планах. Действующая редакция основного прайс-листа публикуется на сайте <http://www.diadoc.ru/pric>. Дополнительные прайс-листы предоставляются по требованию Лицензиата.

1.11. Программный интерфейс (API) Application Programming Interface – интерфейс прикладного программирования, позволяющий провести интеграцию Контур.Диалока с любой учетной системой Лицензиата.

1.12. Ключ разработчика – последовательность символов, буквенно-цифровой код, позволяющий получить доступ к функционалу Контур.Диалока, используя программные методы API.

1.13. Сервисный центр – подразделение Лицензиара или организации/индивидуальный предприниматель, уполномоченные Лицензиаром на основании агентского договора представлять интересы Лицензиара во взаимоотношениях с Лицензиатом.

Диалок. Операторы услуг  
Адрес: 620017, Свердловская обл., г. Екатеринбург, пр-кт. Космонавтов, 56  
Телефон: +7 (343) 272-96-05

### 2. Предмет Лицензионного договора

2.1. Лицензиар предоставляет Лицензиату неисключительное право использования Контур.Диалока на условиях простой (неисключительной) лицензии в пределах, предусмотренных Лицензионным договором, а Лицензиат обязуется принять и оплатить предоставленные права в порядке, установленном Лицензионным договором.

2.2. Если Лицензиату требуется СКЗИ, то Лицензиар обязуется возместить простые (неисключительные) лицензия на право использования СКЗИ на условиях Сублицензионного договора.

2.3. При необходимости Лицензиату могут быть возмездно предоставлены лицензии на иное программное обеспечение, оказаны услуги, а также выполнены работы, предусмотренные прайс-листом Лицензиара.

2.4. Заключение Лицензионного договора рассматривается Сторонами как поручение Оператора персональных данных (Лицензиата по Лицензионному договору) другому лицу, предусмотренное частью 3 статьи 6 Федерального закона Российской Федерации «О персональных данных». При этом Лицензиат поручает Лицензиару осуществление следующих действий (операций) с персональными

# Поддерживаемые входные форматы

- PDF
- BMP:
  - несжатый черно-белый
  - 4- и 8-битный — несжатый Palette
  - 4- и 8-битный — Palette с RLE-сжатием
  - 16-битный — несжатый, несжатый Mask
  - 24-битный — несжатый
  - 32-битный — несжатый, несжатый Mask
- TIFF:
  - черно-белый — несжатый, CCITT3, CCITT3FAX, CCITT4, Packbits, ZIP, LZW
  - серый — несжатый, Packbits, JPEG, ZIP, LZW
  - 24-битный цветной — несжатый, JPEG, ZIP, LZW
  - 1-, 4-, 8-битная палитра — несжатый, Packbits, ZIP, LZW  
(включая многостраничный TIFF)
  - серый — несжатый, Packbits, JPEG, ZIP, LZW

# Поддерживаемые входные форматы

- JPEG:  
серый, цветной
- JPEG 2000:  
серый — Part 1  
цветной — Part 1
- JBIG2:  
черно-белый
- PNG:  
черно-белый, серый, цветной
- GIF:  
черно-белый, LZW  
2-, 3-, 4-, 5-, 6-, 7-, 8-битная палитра, LZW
- PCX, DCX:

# **Поддерживаемые входные форматы**

## **Документы офисных форматов**

- текстовые документы: DOC, DOCX, RTF, HTML, TXT, ODT;
- табличные документы: XLS, XLSX, ODS;
- презентации: PPT, PPTX, ODP.

Для работы с документами офисных форматов требуется пакет LibreOffice или Content AI OpenOffice.

Подробнее в разделе «Установка» --> «Системные требования» Руководства администратора.

# Языки распознавания

Абхазский  
Аварский Гъгъ Гьгь Ллл  
Агульский  
Адыгейский  
Азербайджанский(кириллица)  
Азербайджанский (латиница)  
Аймара  
Албанский  
Алтайский Ö ö Ў ў  
Английский  
Арабский\*  
Армянский (Восточный)  
Армянский (Грабар)  
Армянский (Западный)  
Африкаанс  
Ацтекский  
Баскский  
Башкирский Кк Çç Әә  
Белорусский  
Бемба  
Бенгальский  
Бирманский  
Блэкфут  
Болгарский  
Бретонский  
Буготу|  
Бурятский Ө ө H h  
Валлийский  
Венгерский  
Волоф  
Вьетнамский\*

Гавайский  
Гагаузский  
Галисийский  
Ганда  
Голландский (Бельгия)  
Голландский (Нидерланды)  
Греческий  
Грузинский  
Гуарани  
Гэльский (Шотландия)  
Дакота  
Даргинский  
Датский  
Дунганский  
Дунский  
Зулу  
Иврит\*  
Идиш\*  
Ингушский  
Индонезийский  
Ирландский  
Исландский  
Испанский  
Итальянский  
Кабардино-черкесский  
Казахский  
Калмыцкий  
Каракалпакский  
Карачаево-балкарский  
Каталанский  
Кашубский

Кечуа  
Кикуйю  
Киргизский  
Китайский традиционный\*  
Китайский упрощенный\*  
Конго  
Корейский\*  
Корейский (хангыль)\*  
Корсиканский  
Корякский  
Коса  
Кпелле  
Кроу  
Крымскотатарский  
Кумыкский  
Курдский  
Лакский  
Латинский  
Латышский  
Лезгинский  
Литовский  
Луба  
Лужицкий  
Майя  
Македонский  
Малагасийский  
Малайский (Малайзия)  
Малинке  
Мальтийский  
Мансийский  
Маори

# Языки распознавания

[Марийский](#)

[Минангкабау](#)

Могавк

Монгольский

[Мордовский](#)

[Мяо](#)

Немецкий

Немецкий (Люксембург)

Немецкий (новая орфография)

[Ненецкий](#)

Нивхский

Ногайский

Норвежский

Ньянджа

Оджибве

[Осетинский](#)

Папьяменто

Персидский

Польский

Португальский

Португальский (Бразилия)

Провансальский

Ретороманский

Руанда

Румынский

Румынский (Молдова)

Рунди

[Русский](#)

[Русский \(старая орфография\)](#)

[Русский с ударениями](#)

Саамский

Самоа

[Сапотек](#)

Свази

Себуанский

Селькупский

Сербский(кириллица)

Сербский (латиница)

Словацкий

Словенский

Сомали

[Сото \(Южный\)](#)

Суахили

Сунданский

Табасаранский

Тагальский

Таджикский

Таити

Тайский\*

[Татарский](#)

[Ток-писин](#)

Тонга

Тсвана

[Тувинский](#)

Турецкий

Туркменский(кириллица)

Туркменский (латиница)

[Удмуртский](#)

Узбекский(кириллица)

Узбекский (латиница)

Уйгурский (кириллица)

Уйгурский (латиница)

Украинский

Фарерский

Фиджи

Финский

Французский

Фризский

Фриульский

[Хакасский](#)

Хани

[Хантыйский](#)

Хауса

Хорватский

[Цзинпо](#)

Цыганский

Чаморро

[Чеченский](#)

Чешский

[Чувашский](#)

[Чукотский](#)

Шведский

Шона

[Эвенкийский](#)

[Эвенский](#)

[Эскимосский \(кириллица\)](#)

[Эскимосский \(латиница\)](#)

Эстонский

[Якутский](#)

Японский\*

# Технология ADRT

- Анализирует весь документ как целое, а не отдельные страницы
- Сохраняет точный внешний вид
- Воспроизводит логическую структуру документа при экспорте в Microsoft Word или PDF
  - Колонтитулы
  - Сноски
  - Номера страниц
  - Оглавление со ссылками на соответствующие разделы
  - Подписи к рисункам и диаграммам



# MRC

Создание PDF-документов с ультравысоким сжатием MRC (Mixed Raster Compression), которое позволяет сохранять высокое качество изображения, при этом значительно уменьшать размер PDF-файла документов

***ContentReader Server***  
**А теперь в деталях**





# CONTENTREADER® SERVER

Серверное решение для распознавания, хранения и преобразования файлов в PDF и другие электронные редактируемые форматы. Решение быстро и точно распознает текстовые изображения и переводит их в PDF, PDF/A, DOCX и другие типы файлов, удобные для поиска, просмотра, обмена и хранения.

**Входит в Росреестр, замещает иностранное ПО:**

- ABBYY Recognition Server
- ABBYY FineReader Server
- IRISXtract

Имеет статус «Программное обеспечение относится к сфере искусственного интеллекта» (дает возможность покупателю уменьшить выплаты по налогу на прибыль)



# ContentReader Server

## Схема работы



**Форматы импорта:**  
PDF, BMP, TIFF, JPEG,  
JPEG 2000, JBIG2, PNG, GIF,  
DOC, DOCX, RTF, HTML, TXT,  
ODT, XLS, XLSX, ODS, PPT, PPTX,  
ODP

**Форматы экспорта:**  
PDF, DOCX, XLSX, XML, TXT

**Поддерживаемые ОС:**

РЕД ОС 7.3.2, 8

Astra Linux 1.7.4, 1.8

Ubuntu 22.04

Microsoft® Windows® 11,  
Windows® 10, Windows® 8.1,  
Windows Server® 2022, Windows  
Server® 2019, включая  
развертывание на  
виртуальных машинах

**Более 190 языков  
распознавания**

# Профили обработки

В зависимости от сценария пользователь может выбрать оптимальный профиль обработки документов.

Предопределенные настройки в профилях для типичных сценариев

- Разработаны с использованием алгоритмов машинного обучения
- Позволяют оперативно запустить обработку документов и сэкономить время на ручной настройке комплекса
- Обеспечивают наилучший результат для любого проекта

Сценарий	Конвертация документов с сохранением структуры и форматирования исходного документа (форматы <b>PDF, DOCX, XLSX</b> )	Извлечение текста без сохранения форматирования (форматы <b>XML, TXT</b> )
Приоритет качества	<b>Профиль</b> DocumentConversion_Accuracy	<b>Профиль</b> TextExtraction_Accuracy
Приоритет скорости	<b>Профиль</b> DocumentConversion_Speed	<b>Профиль</b> TextExtraction_Speed

# Производительность CRS 15

Замеры производились на документах PDF без текстового слоя, градации серого, А4, 300 dpi. Содержат таблицы, изображения, текст (договоры, счета-фактуры, товарные накладные).

Процессор Intel Xeon Gold 6238R, 28 core, 2.20GHz.  
x2 (всего на машине 56 ядер, для обработки использовалось от 1 до 48 ядер).

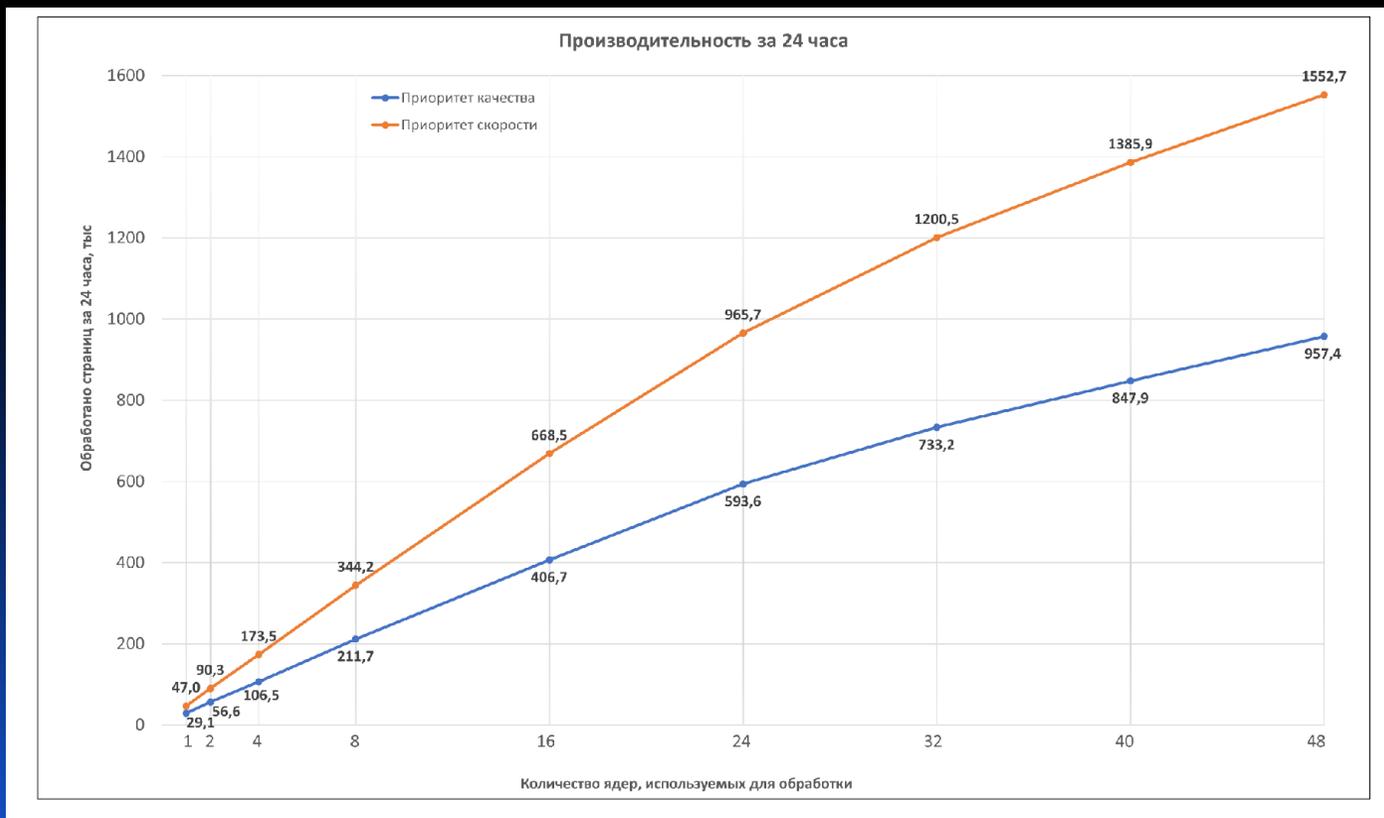
Среднее время обработки одной страницы:

- Профиль с приоритетом качества: **3-4** секунды
- Профиль с приоритетом скорости: **2-3** секунды

За 24 часа непрерывной обработки на 48 ядрах обработали:

- Профиль с приоритетом качества: 957 390 страниц
- Профиль с приоритетом скорости: 1 552 700 страниц

# Производительность CRS 15 в зависимости от количества ядер



# Управление качеством и размером файла при экспорте в PDF

## Параметры:

- Сохранить цветное оформление документа
- Уровень качества изображений в PDF
- Применить алгоритм MRC



# Примеры проектов для демонстрации возможностей API

Языки программирования – C# и Python

Находятся в справке к продукту и содержат:

- Файлы проектов
- Инструкции по сборке и запуску проектов (файлы README.md)
- Пример файла для обработки (в папке Images)

```
def create_job(
    self, params: JobCreationParameters, files: Iterable[FileParameter]
) -> Job:
    """Создать задание на обработку файлов."""

    # Метод принимает multipart/form-data

    # В запросе должен быть один part с именем 'params', содержащий JSON с параметрами создания задания
    request_files = [{"params", (None, json.dumps(params), "application/json")}]

    # В запросе должен быть хотя бы один part с именем 'files', содержащий файл для обработки
    for file in files:
        request_files.append(("files", (file.name, file.stream)))

    response = self._session.post(
        f"{self._base_url}/api/v1/jobs", files=request_files
    )

    # Выбросить исключение в случае неуспешного кода ответа
    response.raise_for_status()
    return response.json()

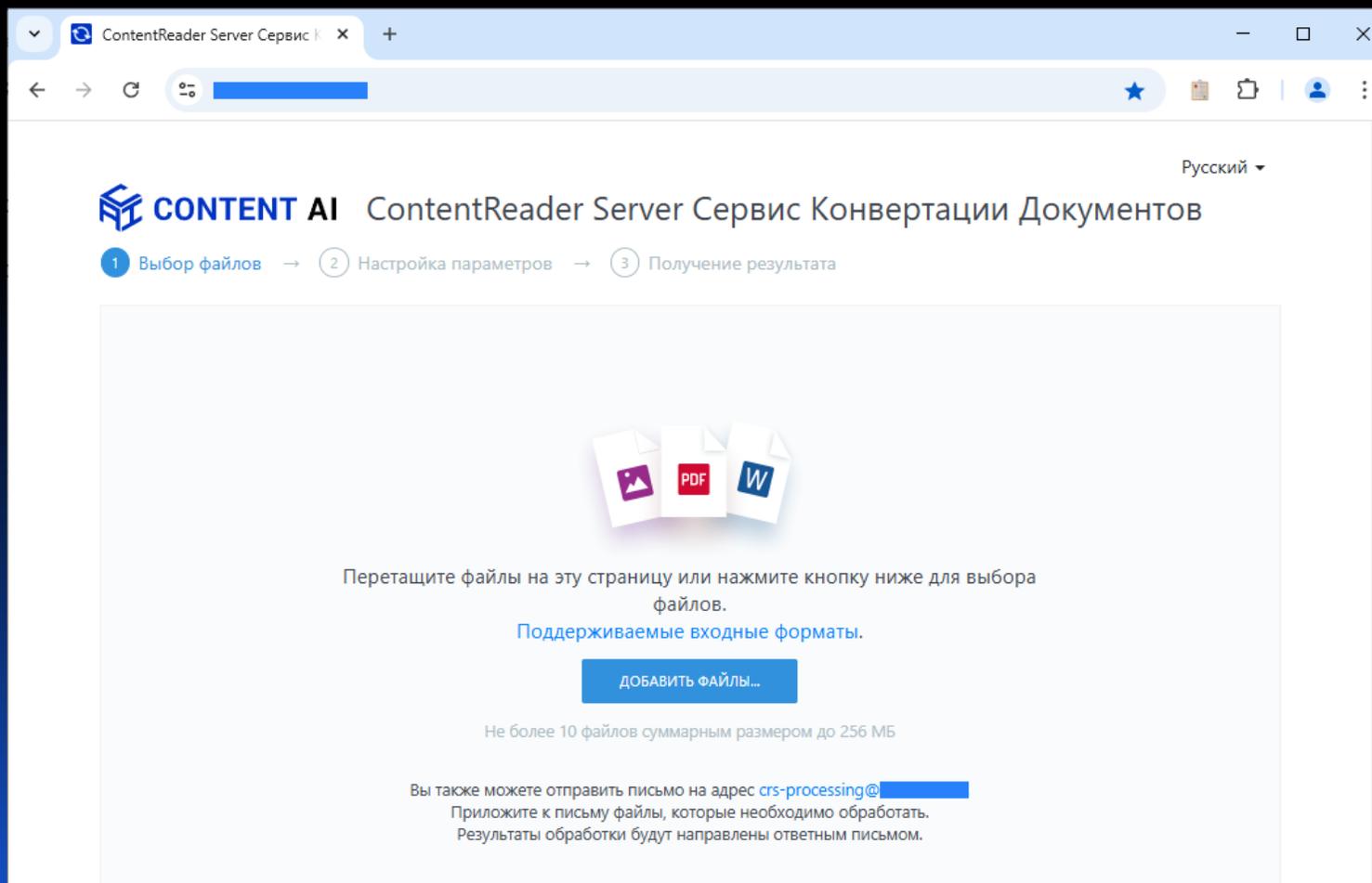
def get_job(self, job_id: str) -> Job:
    """Получить информацию о задании."""

    response = self._session.get(f"{self._base_url}/api/v1/jobs/{job_id}")
    response.raise_for_status()
    return response.json()

def download_job_result(self, job_id: str, result_id: str) -> requests.Response:
    """Получить содержимое файла результата."""

    response = self._session.get(
        f"{self._base_url}/api/v1/jobs/{job_id}/results/{result_id}", stream=True
    )
    try:
        response.raise_for_status()
        return response
    except:
        response.close()
        raise
```

# Пример веб-страницы для конвертации

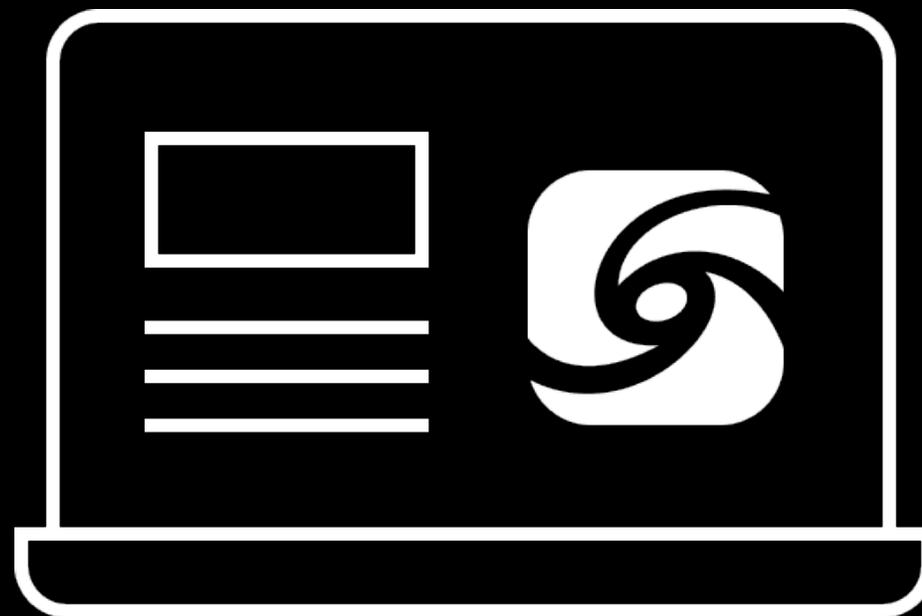




CONTENTREADER® SERVER

## *Демонстрации*

1. Веб-страница администрирования и мониторинга статуса задач
2. Сохранение структуры и форматирования исходного документа (технология ADRT)



***Сценарии***



# Централизованный сервис по оцифровке и конвертации для сотрудников

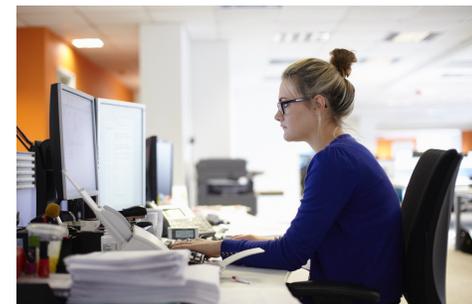
Автоматизированное решение на уровне компании, которое легко масштабируется среди нескольких подразделений, обеспечивая каждого сотрудника универсальным и легким для использования сервисом конвертации документов

Опции использования сервиса:

- Конвертация вложений в электронных письмах
- Обработка файлов из горячей папки
- Сервис для конвертации документов, доступный через браузер

## ПРИМЕРЫ

- Сотрудник отправляет TIFF-файл на специально выделенный почтовый адрес, например, [ocr@company.com](mailto:ocr@company.com) и получает PDF-файл с текстовым слоем в ответном письме
- Сотрудник копирует JPEG-файл в специально созданную папку и получает PDF-файл с текстовым слоем в специально созданной выходной папке
- IT-подразделение компании создает веб-сервис на базе ContentReader Server, доступный в локальной сети компании, где сотрудники могут конвертировать файлы



# Оцифровка архивов и наполнение электронных архивов

Упрощает соблюдение законодательства и локальных корпоративных регламентов, обеспечивая конвертацию большого объема бумажных документов в формат PDF с текстовым слоем



Опции использования сервиса:

- Обработка файлов из горячей папки
- REST API

## ПРИМЕР

- Сотрудник копирует отсканированный документ в горячую папку и PDF-файл с текстовым слоем автоматически попадает в специально созданную выходную папку, откуда затем может быть отправлен в архивную систему



# Регулярная оцифровка и конвертация файлов для дальнейшей обработки

- ContentReader Server может быть использован как предварительный шаг (распознавание и конвертация файлов) для специфических сценариев DLP (data loss prevention), ИИ- анализа, электронного поиска данных (eDiscovery), продвинутых систем поиска и др.
- Возможна интеграция с любой информационной системой



Опции использования сервиса:  
• REST API

## ПРИМЕРЫ

- ContentReader Server распознает отчеты, которые далее будут проанализированы ИИ-системой
- Компания оцифровывает документы с помощью ContentReader Server для импорта в анти-фрод систему или систему электронного поиска данных
- Интегратор или IT-подразделение компании интегрирует ContentReader Server с другим ПО (например, с информационной системой компании), чтобы создать единое решение



# Эффекты от внедрения



**Оптимизация  
документооборота**



**Оптимизация работы  
сотрудников**



**Интеграция через API**



**Старт проекта  
за 15 минут**



**Простота  
использования**



**Возможности поиска**

# Заказчики Content AI



***Вопросы?***

***Протестировать  
продукты Content AI  
на своих документах:***

**[Nikita.Korchagin@softline.com](mailto:Nikita.Korchagin@softline.com)**

