

«Редактор ContentReader PDF и применение ИИ в обработке документов, поиске и извлечении информации от компании Content AI»

Андрей Зеленский

*Руководитель направления партнерских продаж (РФ, СЗФО и страны СНГ)
РФ, Азербайджан, Армения, Беларусь, Казахстан, Кыргызстан, Таджикистан,
Узбекистан*

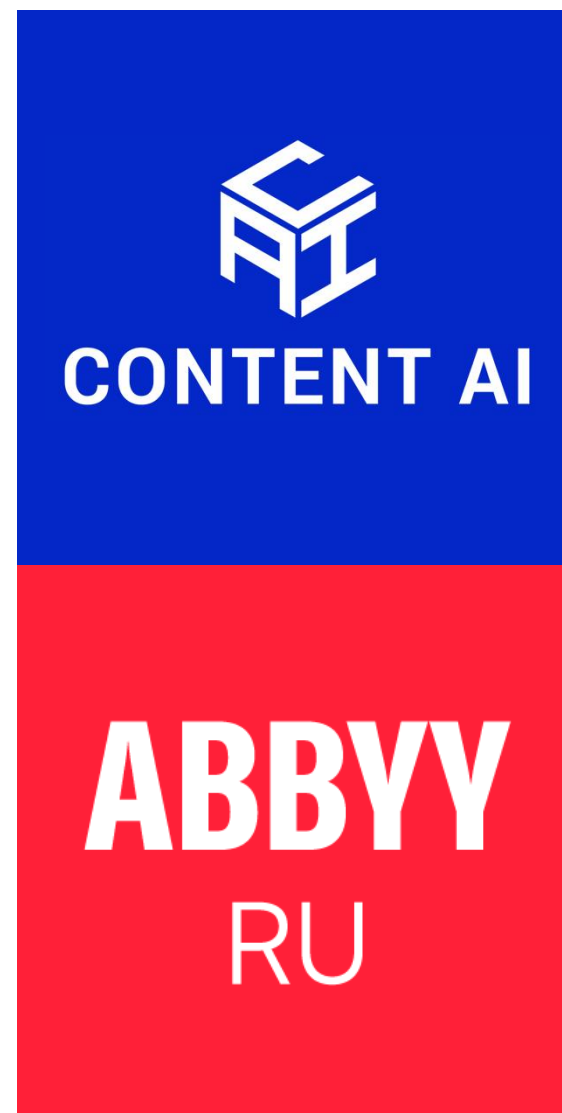
Content AI

Content AI

Content AI (ООО «Контент ИИ») – российская компания, разработчик решений в области интеллектуальной обработки информации. Content AI лицензирует всемирно признанные технологии распознавания текста, классификации документов и обработки естественного языка для крупного бизнеса, государственных организаций и частных клиентов.

Мы предоставляем компаниям и пользователям решения для автоматизации обработки любых документов и данных, востребованные и известные на рынке более 30 лет. Мы помогаем компаниям трансформировать бизнес и делать процессы более эффективными, а людям – добиваться успехов в работе и жизни.

Компания Content AI имеет аккредитацию Минцифры.



Факты о компании



100+
сотрудников в Москве и
Новосибирске



30
лет продукты востребованы на
российском рынке



200+
партнеров на территории
России и СНГ



60%
сотрудников компании
– инженеры и разработчики



Компетенции

Ключевые технологические компетенции Content AI связаны с распознаванием данных в любых типах документов (технологии OCR, ICR, OMR), извлечением информации, обработкой неструктурированного контента (NLP), пониманием и анализом текстов на естественном языке

Обладаем экспертизой в следующих вопросах:



Искусственный интеллект, машинное обучение, технологии обработки естественного языка



Технологические тренды в России и мире



Импортозамещение программного обеспечения



Цифровая трансформация бизнеса



Роботизация бизнес-процессов (RPA)

Ключевые продукты

Интеллектуальная обработка информации



ContentReader PDF

★ теперь и для Linux



ContentCapture

★ уже в реестре



Lingvo by Content AI



Content AI IntelligentSearch

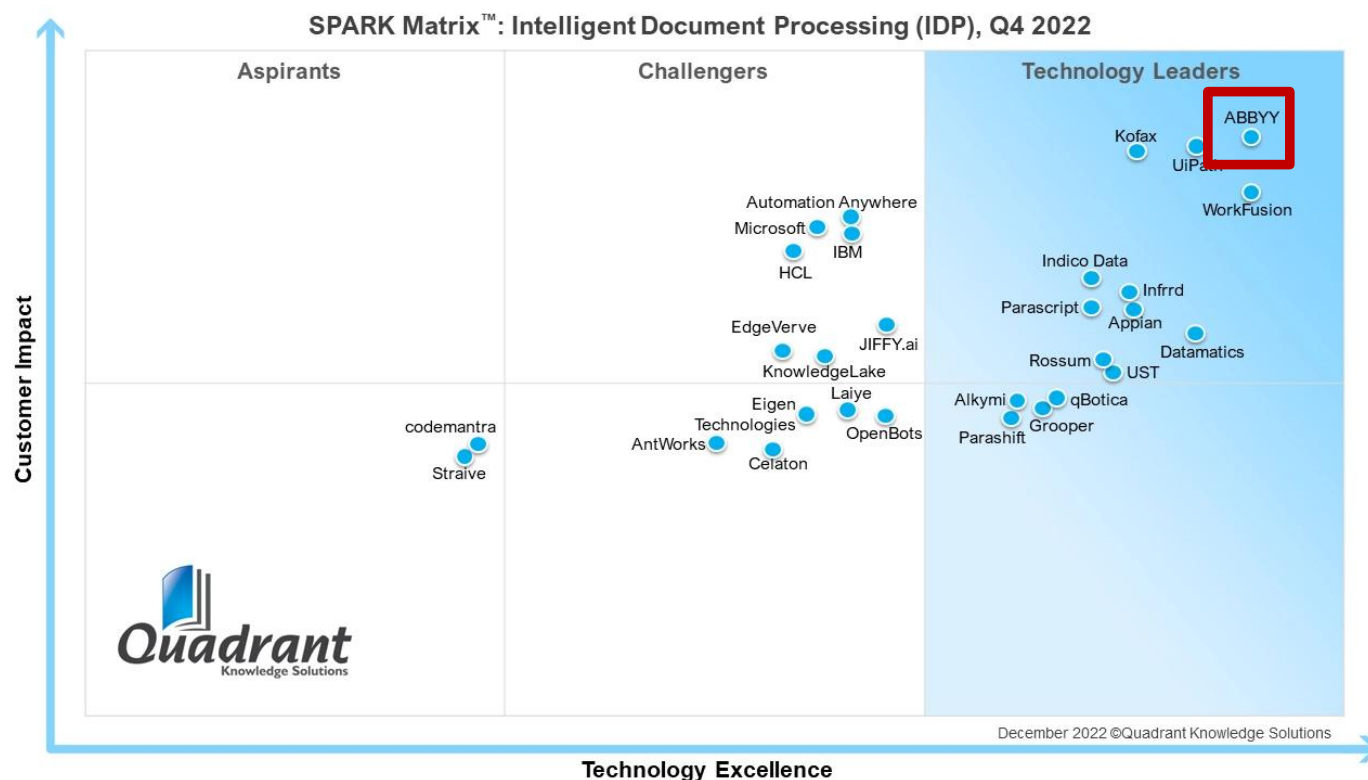


ContentReader Server



ContentReader Engine

Рынок интеллектуальной обработки документов (IDP) в мире в 2022



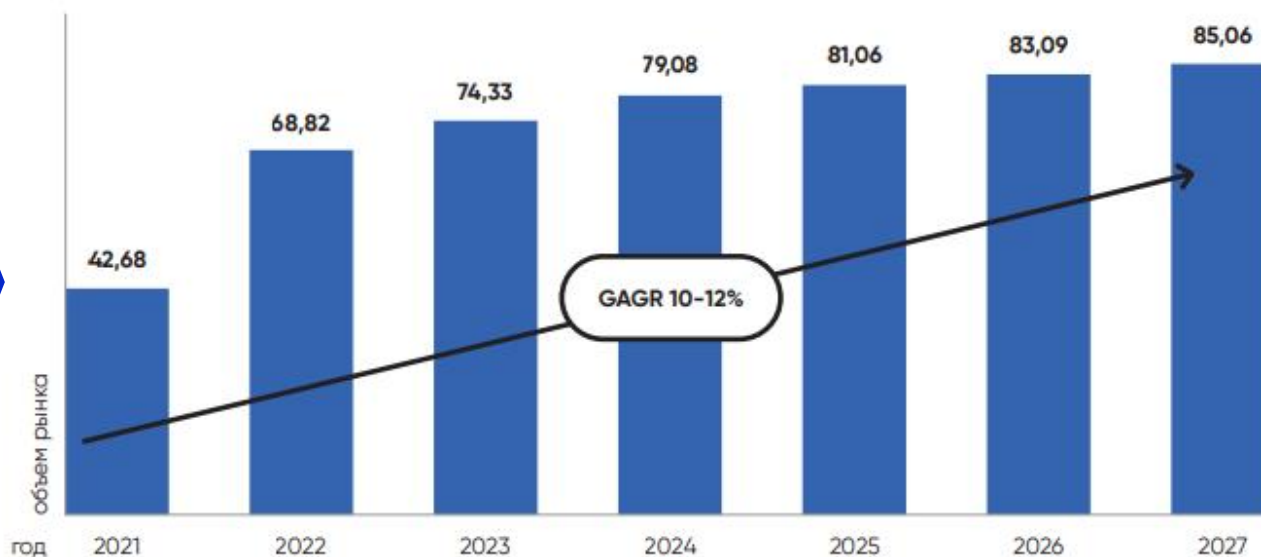
Интеллектуальная обработка документов (IDP) — это технология автоматизации рабочих процессов, которая сканирует, читает, извлекает, классифицирует и упорядочивает важную информацию из больших потоков данных в доступных форматах.

Технология может обрабатывать различные типы документов: бумаги, PDF-файлы, документы Word, электронные таблицы и множество других форматов. Основная функция IDP — извлечение ценной информации из больших наборов данных.

Обзор рынка офисного ПО

По итогам 2021 года объем рынка офисного ПО в России составил 42,5 млрд рублей

Прогноз темпов
роста рынка
офисного ПО,
млрд рублей



Что нам поможет?

Искусственный интеллект - Технологии или Магия?





Adobe Acrobat

ABBYY

FineReader

Nuance Power
PDF

 CONTENT AI

ContentReader PDF



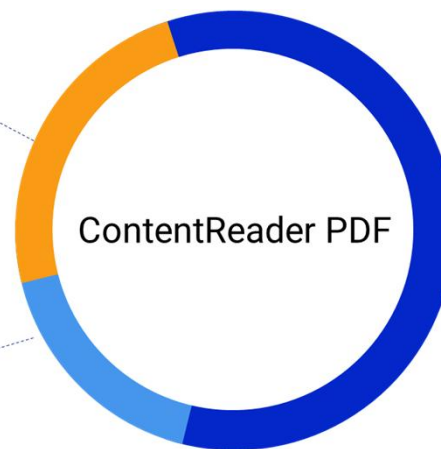
Многофункциональный редактор
для решения любых задач с PDF и
бумажными документами

OCR

Сканирование
Улучшение качества
изображения
Конвертирование
Копирование текста

Сравнение документов

Просмотр различий



Работа с PDF

Просмотр
Создание
Объединение
Редактирование
Комментирование
Заполнение форм
Защита

Ключевые возможности ContentReader PDF

Редактирование PDF	Совместная работа и согласование	Защита и подпись PDF-документов	Открытие и просмотр PDF	Создание PDF	Распознавание, конвертирование PDF	Автоматизация задач по оцифровке и конвертации*	Сравнение документов разного формата*
<ul style="list-style-type: none">• Редактирование текста на уровне абзаца• Редактирование содержания таблиц• Перемещение текстовых блоков, таблиц и изображений• Добавление текста• Операции со страницами• Создание интерактивных форм• Поиск и замена текста	<ul style="list-style-type: none">• Пометки и аннотации• Комментарии и обсуждения• Просмотр всех комментариев на отдельной панели• Поиск, сортировка и фильтр по комментариям	<ul style="list-style-type: none">• Удаление конфиденциальной информации• Удаление скрытых данных• Защита паролем• Добавление цифровой подписи• Добавление изображения подписи	<ul style="list-style-type: none">• ContentReader PDF открывает PDF-документы с на 40% быстрее, чем раньше.• Различные режимы для просмотра страниц, комментариев и метаданных	<ul style="list-style-type: none">• Создание PDF из более чем 25-ти различных форматов• Создание PDF с возможностью поиска• Объединение нескольких документов в один PDF• Уменьшение размера PDF• Разделение PDF по размеру, количеству страниц или закладкам	<ul style="list-style-type: none">• Конвертирование PDF-файлов в редактируемые форматы• OCR-редактор: расширенные инструменты для контроля точности конвертации• Предобработка изображений	<ul style="list-style-type: none">• Автоматизация однотипных или повторяющихся задач по обработке документов• Настройка обработки документов из «Горячей папки»: установите время выполнения задачи или регулярность	<ul style="list-style-type: none">• Быстрое и точное сравнение документов• Акцент на значимых различиях• Экспорт результатов сравнения• Сравнение документов на разных языках

Эффективная работа с PDF-документами

Свободно используйте PDF во всех офисных задачах: конвертируйте, редактируйте и комментируйте любые документы



Adobe Acrobat

ABBYY
FineReader

Nuance Power
PDF

 CONTENT AI

ContentReader PDF

**ВХОДИТ В РЕЕСТР
ОТЕЧЕСТВЕННОГО ПО**

ПОДДЕРЖИВАЕТ LINUX

ЗАМЕЩАЕТ ЗАПАДНЫЕ ПРОДУКТЫ:

- ABBYY® FineReader®
- Adobe® Acrobat®
- Nitro® Pro
- Foxit PhantomPDF®
- Nuance Power PDF
- Soda® PDF и другие

ВЕРСИЯ ДЛЯ LINUX совместима с:

- Astra Linux SE 1.7.0
- РЕД ОС 7.3
- Альт Рабочая станция 10.0
- Альт Рабочая станция К 10.1
- Атлант 2022.02

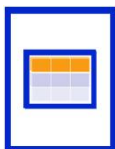
Преимущества решения по сравнению с конкурентами



Наиболее полная функциональность для работы с PDF-документами на Linux



Высокая точность и скорость работы благодаря лучшим AI-технологиям на рынке



Возможность легко копировать таблицы из PDF с сохранением их структуры



Простой и понятный интерфейс, созданный на основе многолетнего опыта и протестированный нашими UX-специалистами



Редактирование текста на уровне абзаца, а не на уровне отдельных слов. Это позволяет работать с PDF, как с обычным текстовым документом



Быстрая и точная конвертация отсканированных документов в редактируемые форматы



Техническая поддержка, подробная справка, возможность доработки решения под запросы клиента

Ключевые возможности ContentReader PDF для Linux

Многофункциональный редактор PDF-документов

Внесение правок
в структуру
и содержание PDF

Совместная
работа
и согласование

Защита и подпись
PDF-документов

просмотр
PDF-документов

Создание
PDF- документов

Поиск по тексту
PDF-документов

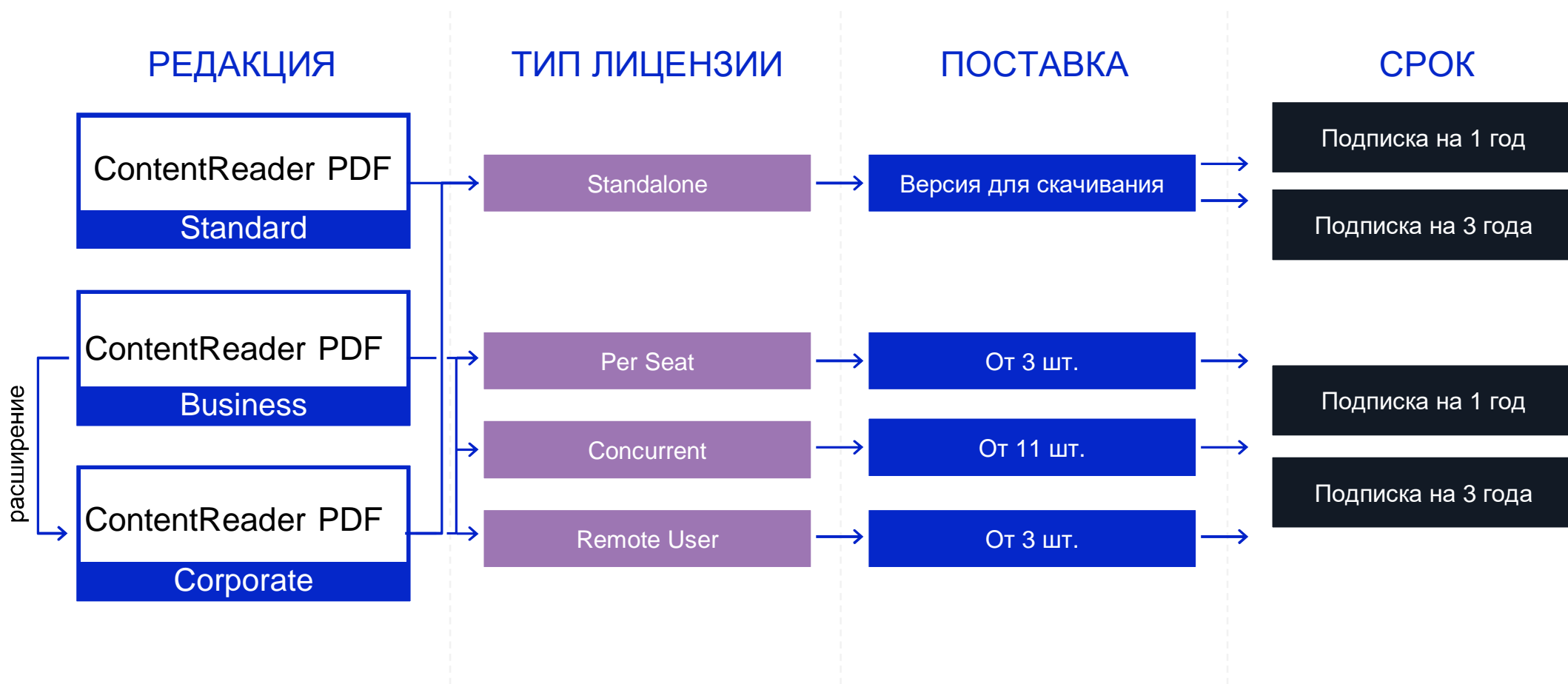
Конвертирование
в редактируемые
форматы

Работа
с PDF-формами

Лицензирование



Схема лицензирования

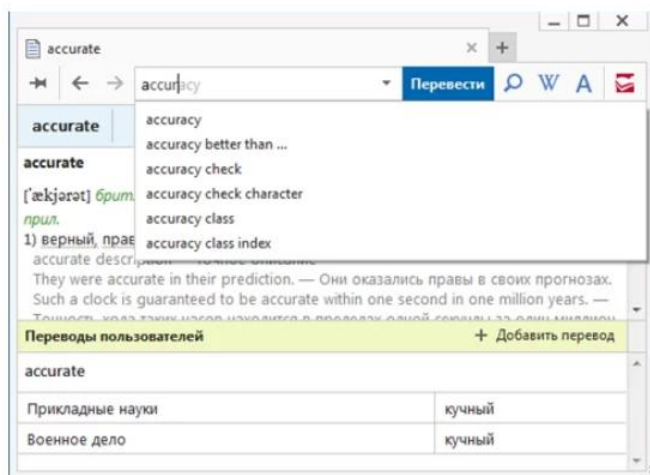


ContentReader PDF Редакции

STANDARD	BUSINESS	CORPORATE
Просмотр и работа с PDF	Просмотр и работа с PDF	Просмотр и работа с PDF
Распознавание и конвертирование	Распознавание и конвертирование	Распознавание и конвертирование
		Автоматическая обработка: 5 000 стр/мес, 2 ядра
		Сравнение документов
* только для домашнего использования!	Лицензирование для организаций	Лицензирование для организаций
Standalone	Per Seat, Concurrent, Remote User	Per Seat, Concurrent, Remote User
от 6 466 ₽	от 11 736 ₽	от 27 346 ₽

Бонус для пользователей: Screenshot Reader

С помощью Lingvo можно



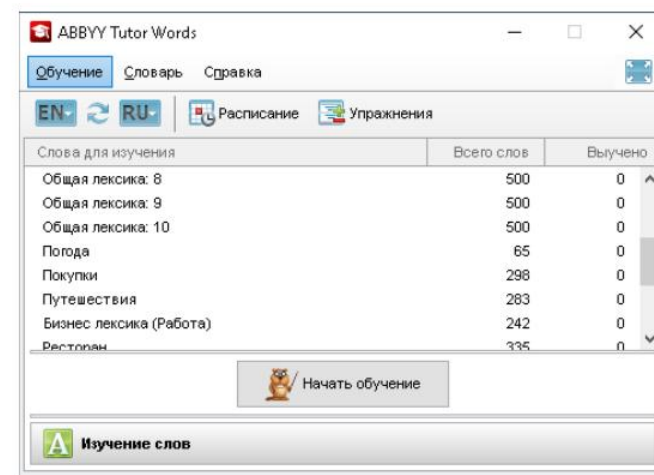
Переводить слова быстро и точно для 12 языков

Переводите быстро и точно с английского, немецкого, французского, испанского, итальянского, португальского, греческого, китайского, украинского, казахского и норвежского на русский язык и обратно



Переводить по наведению в любом приложении

Получите мгновенный перевод, наведя курсор на слово на сайте, картинке, в письме, субтитрах к фильму или PDF-файле. Из окна перевода по наведению можно быстро добавить слово в приложение для запоминания слов.



Запоминать слова с Lingvo Tutor

Получите мгновенный перевод, наведя курсор на слово на сайте, картинке, в письме, субтитрах к фильму или PDF-файле. Из окна перевода по наведению можно быстро добавить слово в приложение для запоминания слов.

Другие решения для интеллектуальной обработки документов

FlexiCapture – ContentCapture
FineReader Server – ContentReader Server
Content AI Intelligent Search

Курс на российских
пользователей

ContentCapture



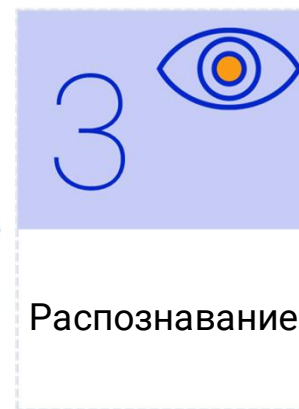
Универсальная платформа для интеллектуальной обработки информации



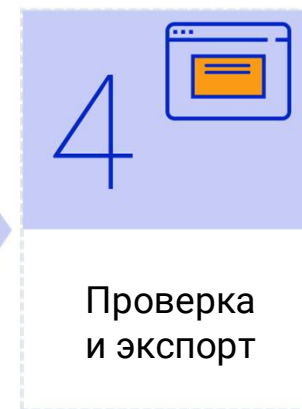
Любые типы документов



Передовые ИИ-технологии



Лучшие качество и скорость



Гибкие настройки

ABBYY
FlexiCapture

IBM

 CONTENT AI



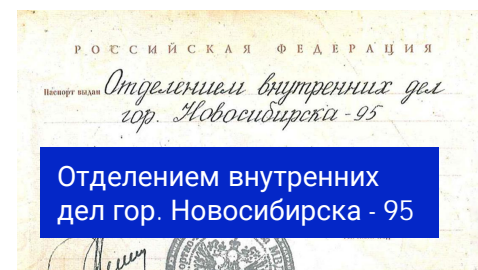
ContentCapture

Основные направления развития

- Linux



- Русский рукописный



- Совместимость с отечественными ПП



- Запись в реестре от 17.05.2023 №17598

Преимущества решения ContentCapture по сравнению с конкурентами



Единый центр ввода для всех типов документов



Интеллектуальные технологии для обработки неструктурированных документов



Повышенная отказоустойчивость и эффективное распределение нагрузки



Широкие возможности удаленной работы с помощью различных веб-приложений



Готовые шаблоны для наиболее востребованных документов: счетов-фактур, товарных накладных и т.д.



Неограниченная масштабируемость

Классификация и извлечение данных из документов любой сложности

Структурированные

CUSTOMER QUESTIONNAIRE

Fill in date:

First and last name:

Occupation:

E-mail:

Company name:

Company Business:

Types of documents to be processed:

Processing volume (pages per month):

How are documents processed in your company at the moment:

Name the manufacturer of the device you plan to use for scanning:

What kind of back-end application or storage do you want to release into?:

Your comments and suggestions:

Слабоструктурированные

EDISON
14155 Bake Pkwy
Irvine, CA 92618
(800) 611-1911

UTILITY BILLING STATEMENT

ACCOUNT NO.	CUSTOMER NAME	SERVICE ADDRESS	BILLING DATE	DUE DATE
000000000	JOHN Q. PUBLIC	1234 Any Street Irvine, CA 92618	01/29/17	02/15/17

WATER READ DATE	To	From	Previous	Current	Consumption	DESCRIPTION	AMOUNT
12/09/16	01/10/17	32	8,544	9,161	617	Previous Balance	287.03
						Adjustments	
						Payments	287.03 CR
						WATER	146.94
						SEWER	115.20
						STATE/COUNTY	8.96
						TAX	2.94
						CITY TAX	0.41
						FUNDS	

WATER CONSUMPTION
This Period: 617 Same Period Last Year: 617

Office Hours: 8:00 a.m. - 5:00 p.m., Monday through Friday. Billing Inquiries: (800) 000-0000

EDISON
14155 Bake Pkwy
Irvine, CA 92618
(800) 611-1911

Make checks payable to: EDISON
PLEASE RETURN THIS PORTION WITH CASHED PAYMENT AND
CHECK ENCLOSED ENVELOPE TO ENSURE PROPER FILING.
CHECK HERE AND COMPLETE BACK OF STUB FOR SURE PAY.

ADDRESS: 1234 Any Street Irvine, CA 92618
RETURN TO: 1234 Any Street Irvine, CA 92618

ACCOUNT NUMBER: 000000000 DUE DATE: 02/15/17 PAY THIS AMOUNT: \$274.45

EDISON
John Q. Public
1234 Any Street
Irvine, CA 92618
(800) 611-1911

Неструктурированные

Regional Analysis

Social Forces
The following data has been published by *Site To Do Business* (www.site-do.com) as estimated demographics information based on the 2010 Census and modeling provided by the US Census Bureau.

Population Profile: The following table summarizes population trends for the Fayetteville, as well as comparative data for the State of North Carolina and the United States.

Region	2000		2010		Growth Rate	Estimated		Growth Rate	Projected		Growth Rate
	2000	2010	2009-2010	2017		2010-2017	2022		2017-2022		
Fayetteville	136,609	346,383	0.65%	382,766	0.65%	396,711	0.72%				
North Carolina	8,049,313	9,555,483	1.71%	10,304,250	1.11%	10,883,629	1.10%				
United States	281,421,906	308,745,538	0.93%	327,314,334	0.85%	341,323,594	0.83%				

Households: The following table summarizes trends in household growth for the region.

Region	2000		2010		Growth Rate	Estimated		Growth Rate	Projected		Growth Rate
	2000	2010	2009-2010	2017		2010-2017	2022		2017-2022		
Fayetteville	118,731	138,963	1.09%	146,557	0.76%	152,119	0.75%				
North Carolina	3,132,013	3,745,355	1.80%	4,031,256	1.06%	4,250,478	1.06%				
United States	105,480,101	116,716,292	1.02%	123,158,887	0.77%	128,069,416	0.79%				

Income Levels: The following table summarizes trends in median household income levels.

Region	2017		Projection		Growth Rate
	2017	2022	2017-2022	2017-2022	
Fayetteville	\$46,023	\$50,341	1.77%		
North Carolina	\$48,918	\$53,833	1.95%		
United States	\$56,124	\$62,316	2.12%		

Conclusions
The Fayetteville area serves as the region's hub for shops, restaurants, services, lodging, health care, employment and entertainment. The local economy is heavily influenced by Fort Bragg and Pope Field and has one of the largest growing defense industries in the nation. The region has displayed an average growth rate in comparison to the state and displays income demographics slightly below the state average. Enhanced by the continued growth of the defense industry and local and national governmental services/employment, the Fayetteville area should remain an increasingly attractive area for investment.

18-485 Worksheet on Metropolitan 4-18-18.rpt Deverick & Associates, Inc. 33

NOTICE TO INDIVIDUAL BENEFICIARY(IES) REGARDING ASSIGNMENT OF INSURANCE POLICY

- The Assignment is an important legal document. Before executing the Assignment, you should read it carefully and seek independent legal advice to ensure that you understand the full consequences of executing the Assignment.
- The Assignment will secure (i) all monies, obligations and liabilities in any currency owing by the Borrower to the Assignee from time to time and at any time actually or contingently, in any capacity, alone or jointly with any other person under the Transaction Documents, (ii) interest on any amount owing by the Borrower (both before and after any demand or judgment), to the date on which the Assignee receives payment, at the rates payable by the Borrower or which would have been payable but for any circumstance which restricts payment and (iii) all fees, commissions, charges, expenses and disbursements (including, without limitation, legal fees) which are of reasonable amount and reasonably incurred by the Assignee in enforcing or attempting to enforce the Assignment (the "Secured Obligations").
- By executing the Assignment, you may become liable instead of the Borrower or together with the Borrower. Your liability under the Assignment in respect of the Secured Obligations will be limited to the amount (if any) specified in Schedule I to the Assignment or the amount of the Proceeds, whichever is less.
- Any money payable to you by the Insurer in respect of the Policy may be used to settle the Secured Obligations if the Borrower fails to pay the Secured Obligations. Any payment in respect of the Policy received by you must be held by you in trust for the Assignee and be paid to the Assignee upon demand.
- The Assignment will be binding as if continuing security on you until the Secured Obligations shall have been paid or satisfied in full and all obligations and liabilities incurred by the Borrower have been duly performed and discharged.
- This Notice is a summary of the salient terms and conditions of the Assignment. Nothing contained in this Notice shall affect or vary the terms and conditions set out in the Assignment. In case of conflict, the terms and conditions of the Assignment shall prevail.

You acknowledge that before executing the Assignment you have been advised to seek independent legal advice on the terms and conditions of the Assignment. You also acknowledge that you have been provided with a copy of the Assignment.

Expressions used in this Notice shall have the same meanings as defined in the Assignment.

Optical Character Recognition OMR Intelligent Character Recognition BCR Natural Language Processing

Технологии

OCR – Optical character recognition

202 languages, including:

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

αα

HANDPRINT

HANDPRINT

HANDPRINT

H A N D P R I N T

H A N D P R I N T

H A N D P R I N T

H A N D P R I N T

H A N D P R I N T

ICR – Intelligent character recognition

136 languages

• 1 2 3 4 5 6 7 8 9

1 2 3 4 5 6 7 8 9 0
1 2 3 4 5 6 7

OMR – Optical mark recognition

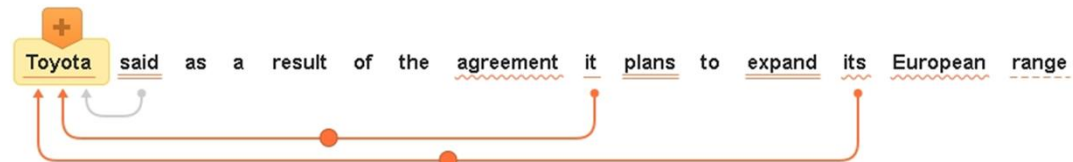


BCR – Barcode recognition



NLP – Natural Language Processing

CONTENT AI



Обработка данных из врачебных направлений

Медицина

ЗАДАЧИ

- Снизить нагрузку на сотрудников лаборатории и избавить их от непрофильной деятельности;
- Ускорить прием пациентов;
- Сократить ошибки при вводе данных в медицинскую ИС.

РЕЗУЛЬТАТ

- **В 4 раза** сократилась занятость персонала на перепечатку данных из документов;
- Пациенты **быстрее** получают результаты;
- Минимизация ошибок при вводе данных.



Кейс «ТЭК-Торг»: автоматизация проверки закупочной документации

Задача

- Ускорить подачу заявки на участие в тендерах для клиентов
- Уменьшить количество ошибок
- Привлечь новых клиентов за счет более удобного сервиса

Результаты:

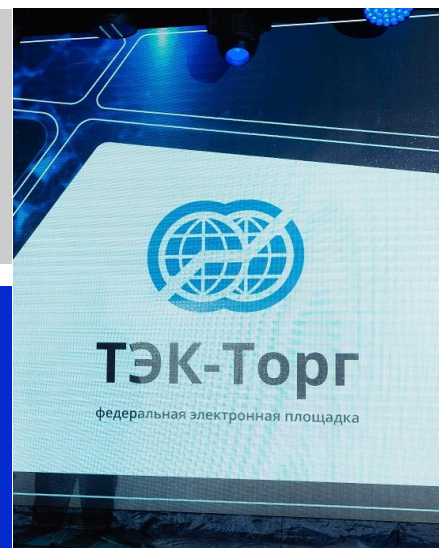
В 10 раз

сократилось время проверки закупки

Менее 1 мин

уходит на обработку одного документа

Клиенты «ТЭК-Торг» быстрее организуют закупки и получают заказы от поставщиков



ContentReader Server

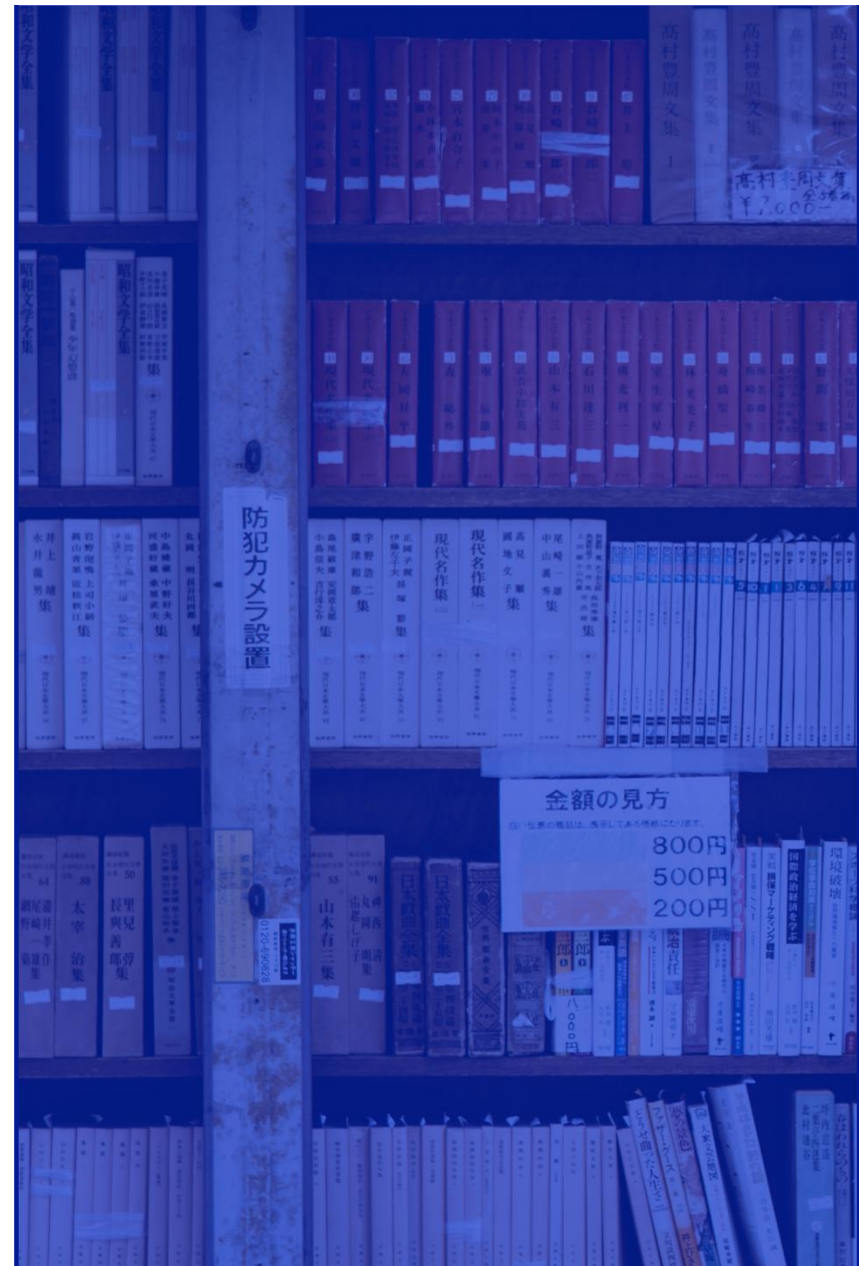
ex. FineReader Server



Серверное решение для распознавания и конвертации больших объемов документов

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, Microsoft Word или другие типы файлов, удобные для поиска, просмотра, обмена и хранения.

 **CONTENT AI**



С помощью ContentReader Server можно создать архивы материалов, оцифровать входящую документацию

Задачи

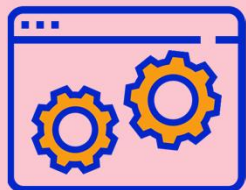
- § Создание архива материалов в организации с удаленным доступом
- § Снижение затрат на бумагу и помещения для ее хранения
- § Упрощение поиска и доступа к нужным материалам

Решение



Этап 1.

Сотрудник сканирует материалы



Этап 2.

ContentReader Server осуществляет полнотекстовое распознавание



Этап 3.

При необходимости сотрудник проверяет полученный текст



Этап 4.

Распознанные материалы сохраняются в архив в цифровом формате, удобном для поиска и редактирования

Создание электронного архива для хранения документов

Задача

Перейти от бумажного архива к электронному
Обеспечить быстрый доступ к документам в случае проверок
Сделать работу с документами более простой и удобной

Результаты:

35 000

документов обрабатываются в системе ежемесячно



С 10 до 3 дней

Уменьшились сроки обработки документов

Качество обработки входящих документов повысилось: если подписанные документы возвращаются в потрепанном виде, их можно проверить по штрихкоду

Coveo
ABBYY
Intelligent
Search
Daxtra

 CONTENT AI

Content AI Intelligent Search



Готовое решение для создания
корпоративного поискового портала



Coveo
ABBYY
Intelligent
Search
Daxtra

 CONTENT AI

Content AI Intelligent Search

**ВХОДИТ В РЕЕСТР
ОТЕЧЕСТВЕННОГО ПО**

ПОДДЕРЖИВАЕТ LINUX

ЗАМЕЩАЕТ ЗАПАДНЫЕ ПРОДУКТЫ:

- 3i Data Plexus
- Attivio
- Coveo
- Daxtra
- dtSearch
- EPAM InfoNgen
- IBM Watson Discovery
- LucidWorks Fusion
- Microfocus IDOL
- Mindbreeze
- Nuix
- Sinequa

ВЕРСИЯ ДЛЯ LINUX совместима с:

Операционная система Linux:

- Ubuntu 20.04;
- RED OS 7.3 Murom;
- Astra Linux 2.12.42 (Orel, Common Edition);
- ALT Server 9.1.

Преимущества решения Intelligent Search по сравнению с конкурентами



Единая точка доступа ко всем корпоративным источникам



Поиск по точному совпадению и по смыслу (семантический)



Встроенная технология OCR для поиска по pdf файлам



Быстрое развертывание: первый результат за 30 мин.



Поддержка отечественных ОС на Linux: Astra Linux, RED OS, ALT Linux



Готовые коннекторы для распространенных корпоративных источников

Взгляд в будущее IDP

- Важно говорить не просто об импортозамещении, но об опережающем развитии
- Бизнес транслирует запрос на “повышение интеллектуальности” продуктов
- Для удержания лидирующих позиций в индустрии необходимо ориентироваться на передовые ИИ технологии

- Разрабатываем решения с применением передовых ИИ-технологии
- Уже используем алгоритмы ИИ во всех наших продуктах
- Расширяем использование ИИ на всех этапах работы с продуктом

552 млрд

совокупный объем рынка ИИ на основе данных о выручке компаний

Источник: МФТИ

Более 30%

организаций планируют внедрить ИИ в перспективе ближайших 3 лет

Национальный центр развития искусственного интеллекта при Правительстве Российской Федерации

В 5-6 раз

увеличивается скорость и качество выполняемых работ в отдельных компаниях поле внедрения ИИ-решений

Национальный центр развития искусственного интеллекта при Правительстве Российской Федерации

Эффект от внедрения



Сокращение времени принятия решений

Быстрая и точная обработка данных ускоряет принятие управленческих решений.



Снижение издержек на обработку документов

Автоматическая классификация, распознавание и экспорт данных из любых источников и типов документов в ИС.



Контроль на всех этапах и уровнях

Сокращение ошибок при вводе данных, связанных с человеческим фактором.

Спасибо

Контакты

Андрей Зеленский.

Руководитель направления партнерских продаж (РФ, СЗФО и страны СНГ)

РФ, Азербайджан, Армения, Беларусь, Казахстан, Кыргызстан, Таджикистан, Узбекистан

Content AI

E-mail: andrew.zelenskiy@contentai.ru

Адрес: Россия, г. Москва, ул. Отрадная, д. 2Б, стр. 6

E-mail: office@contentai.ru

Web: www.ContentAI.ru