

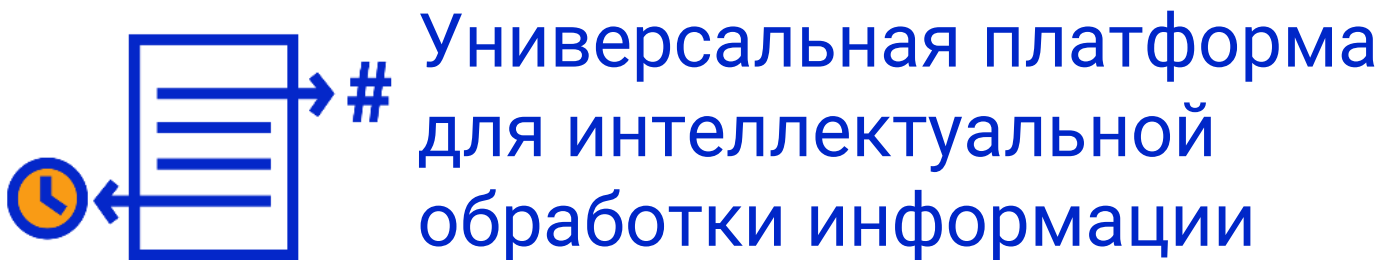
ContentCapture 14

Новости о продукте

Злата Баяндина,
Старший менеджер продукта

Ольга Гусева,
Специалист по тестированию

ContentCapture



ContentCapture – универсальная платформа для интеллектуальной обработки информации из любых типов документов: отсканированных бумаг, фотографий, электронных документов, текстов писем и вложений.

Решение распознает, классифицирует документы, извлекает данные, проверяет их корректность и передает в корпоративные информационные системы.



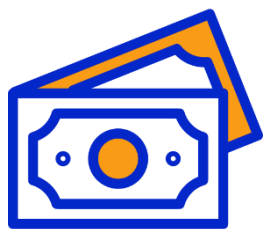
Входит в реестр
отечественного
ПО и замещает
иностранное ПО:

- ABBYY
FlexiCapture
- KofaxCapture
- IBM Datacap

ContentCapture - решение на базе ИИ

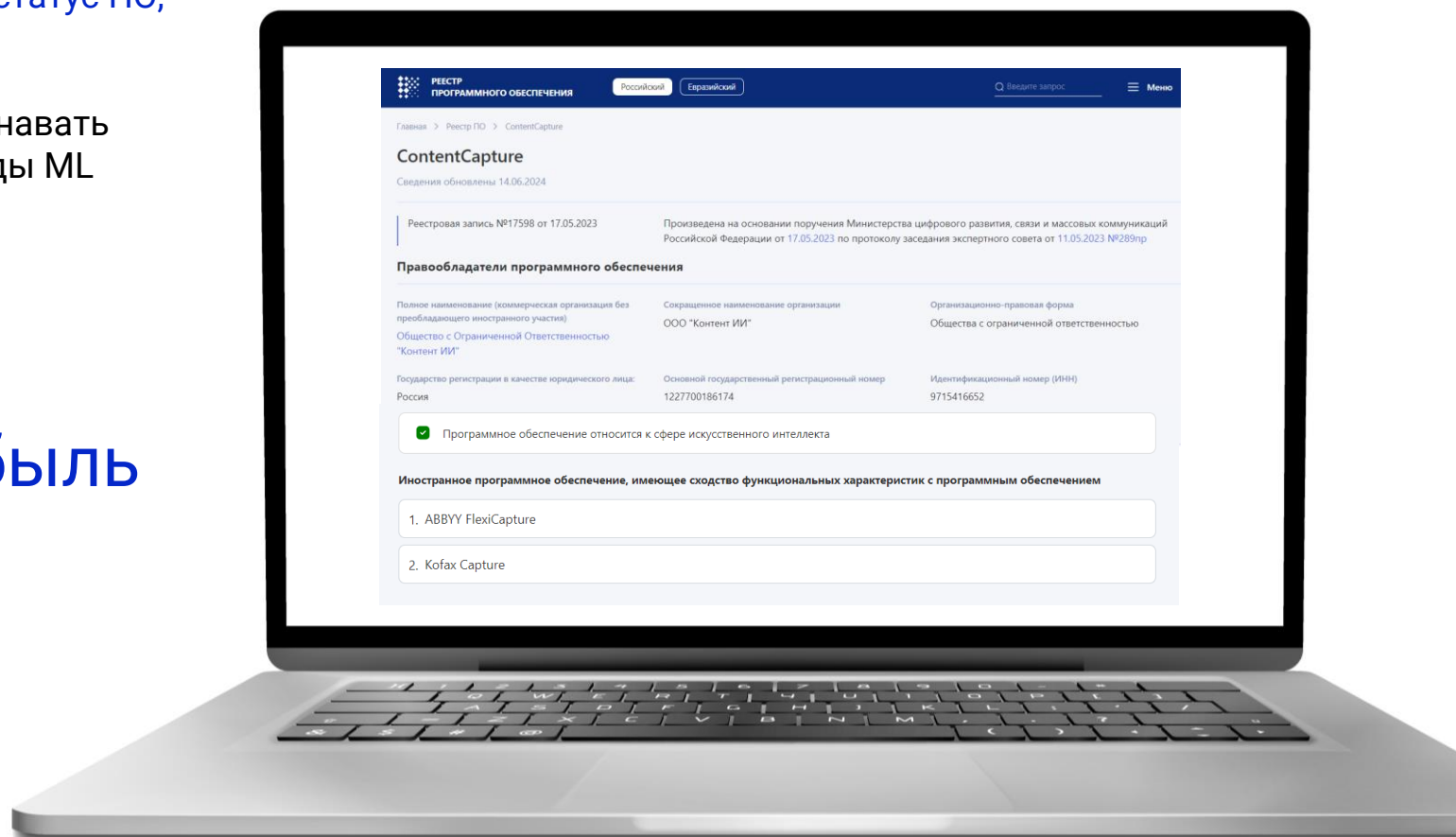
В реестре отечественного ContentCapture получила статус ПО, относящегося к сфере искусственного интеллекта.

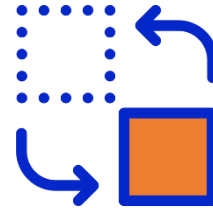
Нейросети позволяют с высокой точностью распознавать данные в любых документах и изображениях, методы ML используются для обнаружения и классификации



**Возможность
снизить выплаты
по налогу на прибыль**

Расходы на ПО, которое включено в единый реестр российских программ для ЭВМ и баз данных и которое относится к сфере искусственного интеллекта, можно учитывать для целей налога на прибыль с коэффициентом 1,5.



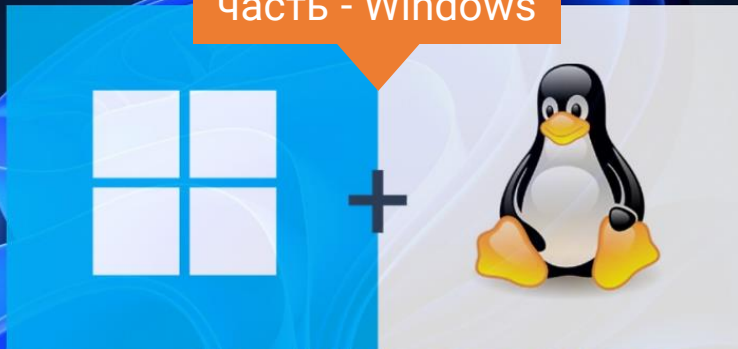


Стадии в процессе перехода с импортного на отечественное ПО

Серверная и клиентская часть на Windows



Серверная часть – Linux
Клиентская часть - Windows



Серверная и клиентская часть на Linux



Схема работы решения



Счет, счет-фактура,
товарная накладная
и др.

ИМПОРТ
ИЗОБРАЖЕНИЙ И
ДОКУМЕНТОВ



- Классификация
- Комплектность
- Извлечение данных
- Проверка и сравнение

ОБРАБОТКА
ДОКУМЕНТОВ



Проверка
и контроль
важных данных

ВЕРИФИКАЦИЯ
ДАННЫХ



ЭКСПОРТ
В ЦЕЛЕВЫЕ
СИСТЕМЫ

2024: новые функциональные возможности

FastML для обучения

Появилась возможность быстрого добавления новых типов документов через разметку – достаточно 5-10 документов

Русский рукописный

Доработан процесс распознавания рукописного – можно работать с буквоцифрами, сложным почерком и многострочными полями

Печати и подписи

Добавлена детекция таких элементов, как печать и подпись на документе. В Layout Studio появились новые типы блоков

Установка и настройка

Обновили инструменты для установки, настройки и администрирования комплекса – инсталлятор для Windows, пакеты deb, rpm для Linux и др.

2024: другие изменения в ContentCapture 14

ContentCapture 14.1:

- Станция настройки проектов на Linux
- Станция веб-верификации
- Импорт/экспорт по протоколу SMB в «толстых» станциях (пути в нотации \\servername\path)

ContentCapture 14.2:

- Монитор станции обработки
- Аутентификация при помощи технологии единого входа (SSO, Single Sign-On)
- Добавлены: утилита fcadmintools, экспорт в формат XML, экспорт в БД PostgreSQL
- Изменился номер версии БД – 144
- API несовместимо с предыдущими версиями
(другие адреса end point'ов, удалены некоторые методы)

Добавление новых типов документов без разработки гибких описаний

Сценарии применения



Демонстрация

Быстрая подготовка проекта для демо платформы заказчику. Возможность показать, как работает СС на документах пользователя* в реальном времени



Конфиденциальные документы

Возможность настройки платформы на стороне заказчика на этапе самостоятельного тестирования платформы.



Разработка гибких описаний

Экономия времени и ресурсов на создание или улучшение гибких описаний для заказчика – инструмент можно использовать для создания простых шаблонов



Дообучение текущей системы

В случаях, когда в процессе работы появляются новые документы или новые поля на существующих документах, и их необходимо оперативно добавить

*важно учитывать границы применимости

Добавление новых типов документов без разработки гибких описаний

Схема работы

Для каждого нового типа документов:



1. Выбрать от пяти документов этого типа. Документы должны быть хорошего качества (без дефектов, шума, мусора и т.д.)
2. Разметить вручную нужные поля
3. Запустить процесс обучения
4. Протестировать качество на документах, не использовавшихся в обучении

Рекомендация по делению выборки: обучающая – 80/70%, тестовая – 20/30%.

Обученное гибкое описание можно экспортировать в Layout Studio.

Границы применимости технологии добавления новых типов документов без разработки гибких описаний

МОЖНО ОБУЧАТЬ

1. Тестовые поля
2. Штрих-коды
3. Метки
4. Повторяющиеся группы

НЕЛЬЗЯ ОБУЧАТЬ

1. Новые объекты – печати и подписи
2. Рукописный текст
3. Группы меток

Демонстрация

Технология добавления новых типов документов без разработки гибких описаний

Распознавание русского рукописного

Сценарии применения



Обработка кадровых документов

Извлечение данных из рукописных полей паспортов, трудовых книжек, военных билетов, сертификатов, дипломов и пр.



Обслуживание клиентов

Обработка заявлений, анкет и прочих документов от клиентов в сервисных компаниях, включая государственные организации



Образование

Тесты, экзамены и другие массовые проверочные работы, которые включают в себя рукописную часть



Бэк-офис

Извлечение данных из внутренних документов на производстве, из логистических документов, из архивов и пр.

Демонстрация

Распознавание
русского рукописного текста

Детекция подписей и печатей

Сценарии применения



Обработка первички

При получении от контрагента таких документов как акт, счет-фактура, ТОРГ-12 и пр. необходимо убедиться в наличии печатей и подписей



Открытие счета/кредитование

В сценариях открытия счета юр.и физ. лиц, потребительского кредитования и пр., где есть потребность междокументального сравнения подписей



Логистические документы

Перевозочные документы грузов, ТТН, доверенность на получение товара, договор между контрагентами и проч.



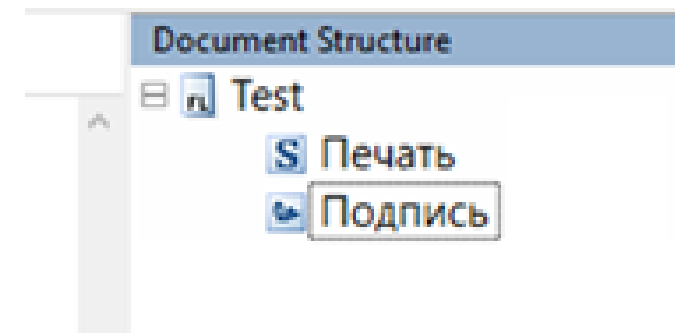
Прочие документы

Для целей проверки подписей и печатей в рамках любого процесса обработки документов и оперативного их отклонения, если элементы не обнаружены

Детекция подписей и печатей

Функциональность

- Появились новые типы блоков, позволяющие найти блоки этих типов на документах и использовать их при разработке гибких шаблонов:
 - подпись
 - печать
- Для блока типа «печать» определяется вид печати:
 - круглая
 - треугольная
 - другая



Демонстрация

Подписи и печати

Улучшения, связанные с установкой и настройкой платформы



Установка



Переход с CC/FC12 на CC14



Изменение в API



Обновление ContentCapture 14

Новые форматы поставки ContentCapture 14

Linux:

- докер-образы
- deb-пакеты **новое**
- rpm-пакеты **новое**

Windows

- **инсталляционный пакет (автоматизирует большинство действий пользователя по установке и настройке продукта, существенно переработан по сравнению с ContentCapture 12)**

CONTENT AI Решения ▾ Продукты ▾ Партнеры ▾ Поддержка О компании ▾ [Консультация](#) [Магазин](#)

ContentCapture 14

Номер версии: 14.2

- [ContentCapture 14.2. Примечания к выпуску](#)
- [Руководство системного администратора](#)
- [Руководство по настройке проектов](#)
- [Справка \(.chm\)](#)

Выберите комплектацию для скачивания:

- [Windows](#) →
- [Докеры](#) →
- [DEB-пакеты](#) →
- [RPM-пакеты](#) →

Изменения в публичном API

Как адаптироваться к изменениям в публичном API:

1. Получить из Сервера приложений формальное описание интерфейса сервиса (wsdl) по следующим ссылкам:

[/Main/wsdl](#)

[/AuthAPI/v2/Soap/wsdl](#)

[/API/v2/Soap/wsdl](#)

[/DatasetAPI/v2/Soap/wsdl](#)

[/IdentityProviders/wsdl](#)

указав адрес Сервера Приложений (например, [https://\[IP|DNS имя\]/ContentCapture/Server](https://[IP|DNS имя]/ContentCapture/Server))

2. Регенерировать клиентский модуль

Изменения API

- В метод аутентификации добавлен ClientID
 - Раньше формировался автоматически `{LoginStationURL}/{TenantName}/AccessToken/Saml``, теперь настраивается.
- Удалены методы
 - AddUserIfNotExist
 - CreateRequest
 - DeleteRequest
 - GetRequestStatus
 - SetRequestStatus
 - GetRequests
 - GetCurrentUserAuthTicket
 - DuplicateProject
 - GetWhitelPList
 - SetWhitelPList
- Изменены описания интерфейсов:
 - С `targetNamespace="urn:http://www.abbyy.com/FlexiCapture"`
 - На `targetNamespace="urn:https://www.contentai.ru/ContentCapture"`
- Добавление версионирования
 - Старый URL: `{BaseURL}/API/Soap``
 - Новый URL: `{BaseURL}/API/v2/Soap``

Переход с предыдущих версий на ContentCapture 14 на Windows

Не рекомендуем устанавливать на одну машину ContentCapture 12 (или FlexiCapture 12) и ContentCapture 14.

Рекомендуемая последовательность шагов миграции:

- 1) Сделать резервные копии БД и файлового хранилища
- 2) Сделать локальные резервные копии всех проектов
- 3) Удалить старую версию ContentCapture 12
- 4) Установить новую версию ContentCapture 14
- 5) Создать базу данных при установке сервера приложений или вручную с помощью утилиты AppServerTools
- 6) Загрузить проекты на сервер

Замечание. Если используются кастомные ключи реестра, то нужно перенести их.

Подробнее см. в документации к продукту.

Обновление с предыдущих релизов ContentCapture 14 на Linux

1. Обновить БД до версии **144** с помощью утилиты AppServerTools.
2. Обновить проекты с помощью открывания каждого проекта на станции настройки проектов или с помощью утилиты FCAdminTools.
3. (В зависимости от используемого способа развертывания.) Обновить докер-контейнеры до версии ContentCapture 14.2 или установить компоненты ContentCapture 14.2 из deb- или rpm-пакетов.

Подробнее см. в документации к продукту.

Дорожная карта ContentCapture 14

В ближайших релизах ContentCapture 14 запланирована следующая функциональность:

- модуль NLP на Linux
- веб-станция сканирования на Linux (с ограниченным функционалом)
- модуль Layout Studio на Linux
- поддержка хранилищ S3 в качестве файлового хранилища
- доработки удобства конфигурирования и диагностики на Linux
- улучшение производительности на Linux и для сценария фронт-офиса
- улучшение качества обработки российских идентификационных документов (российский паспорт и др.)
- улучшение качества распознавания русского рукописного
- отчеты в консоли администрирования и мониторинга
- небольшие доработки UX толстых станций
- валидация подписей

Спасибо! Вопросы?

Контакты

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

E-mail: office@contentai.ru

Web: contentAI.ru



@content_ai

