



Описание шаблона технического решения «Интеллектуальная обработка документов»

Версия 1.0

Содержание

Основные понятия	4
Схема обработки документов.....	5
Обработка поступивших документов	6
Проверка полноты комплекта	7
Верификация документов.....	7
Особенности ввода документов в систему.....	8
Документы со сканера.....	8
Письма и документы с электронной почты	9
Возврат документов от контрагентов	9
Связи документов.....	10
Присвоение номера документам	11
Архитектура решения.....	12
Установка и настройка решения	14
Контур для разработки и тестирования.....	14
Установка решения и внешнего окружения	14
Настройка решения и внешнего окружения	15
Адаптация решения под бизнес-задачи организации	17
Продуктивный контур	19
Установка решения и внешнего окружения	19
Настройка решения и внешнего окружения	19
Обновление продуктивной системы	20
Приложение 1. Команды командной строки для настройки решения	21
Настройка DirectumRX для работы с DIRECTUM Ario	21
Создание классификатора	22
Импорт модели классификации.....	22
Обучение классификатора.....	23
Просмотр моделей, используемых в классификаторе.....	24
Публикация модели классификации	25
Экспорт модели классификации	25
Приложение 2. Настройка службы ввода документов	27
Настройка ввода документов из файловой системы.....	28

Настройка ввода документов с электронной почты	28
Папки для лог-файлов и временных файлов службы DCS	29
Папки для временных файлов DirectumRX.....	29
Приложение 3. Пример адаптации решения	31
Постановка задачи	31
Этап 1. Создание типов документов	31
Этап 2. Классификация документов и извлечение фактов	32
Этап 3. Режим верификации	34

Введение

Работа с документами, поступающими в организацию, начинается с первичной обработки. Подход, который используется в системе, построенной на основе решения «Интеллектуальная обработка документов», позволяет освободить сотрудников от рутинных действий, экономить время и сократить количество ошибок на этом этапе. Решение реализует «умную» обработку входящего потока:

- захватывает документы со сканера и электронной почты, делит их на комплекты;
- распознает текст, классифицирует документы по типам, извлекает факты из текста с помощью сервисов интеллектуальной обработки [DIRECTUM Ario](#);
- заносит документы в DirectumRX и заполняет карточки фактами, извлеченными из текста.

После занесения документов в систему сотруднику остается только верифицировать их: проверить, что все документы комплектов занесены в полном объеме и у них корректно заполнены карточки. При этом распознанные факты подсвечиваются в тексте и в полях карточки, упрощая проверку.



На текущий момент реализована интеллектуальная обработка входящих писем, входящих счетов, накладных, актов, счетов-фактур и универсальных передаточных документов. При этом набор обрабатываемых документов может быть расширен под потребности конкретного заказчика.

Возможности решения предназначены для сотрудников, которые занимаются первоначальной обработкой и регистрацией документов: делопроизводителей, сотрудников бухгалтерии и договорного отдела.

Особенности использования

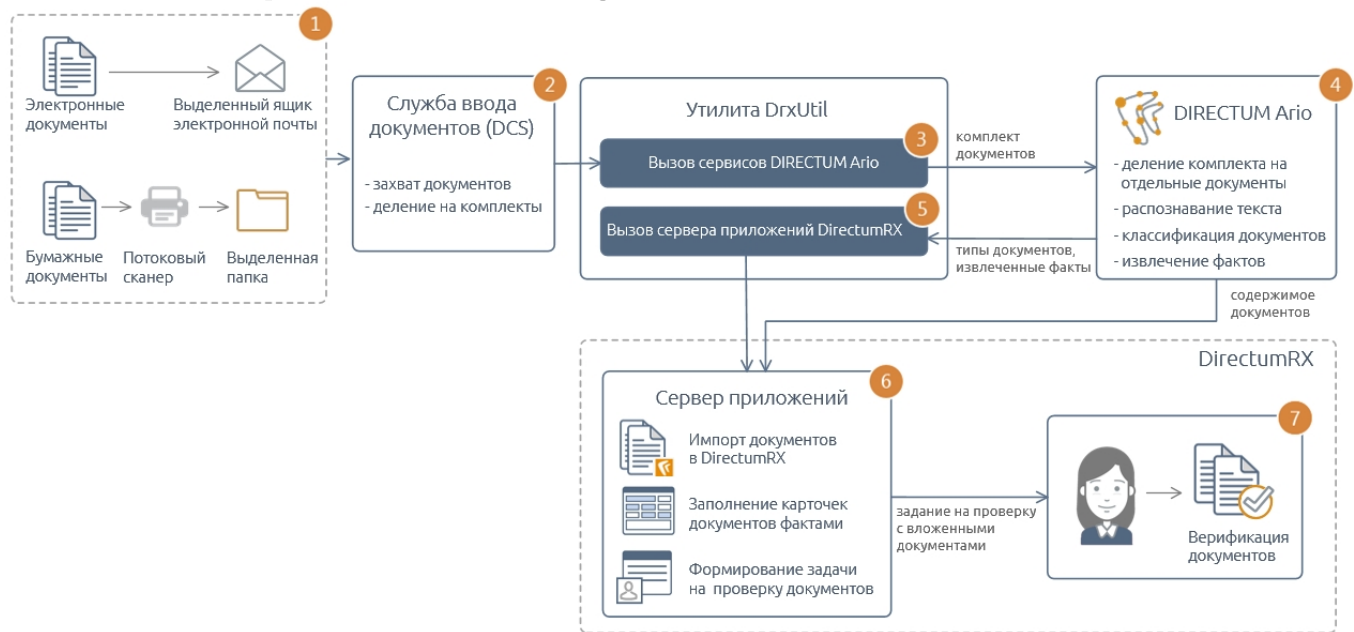
Обратите внимание, что решение «Интеллектуальная обработка документов» на текущий момент имеет статус **шаблона технического решения**. Шаблон предполагает значительную доработку под проект внедрения и потребности организации, в отличие от готового решения, которое используется без внесения крупных изменений в разработку. Один из возможных вариантов расширения функциональности – добавление обработки договоров и дополнительных соглашений.

Основные понятия

Комплект документов

Несколько документов, поступивших от контрагента одним пакетом. Комплект состоит из ведущего документа и приложений. Например, в комплекте, подтверждающем совершение хозяйственной операции, ведущий документ – акт или накладная, а приложение – счет-фактура.

Схема обработки документов



1. Поступление комплекта документов:

- электронные документы поступают на выделенный ящик электронной почты;
- бумажные документы сотрудники сканируют, разделяя комплекты документов белыми листами. Отсканированные документы попадают в выделенную папку.

2. Захват документов. Служба ввода (DCS):

- забирает поступившие документы из выделенного ящика электронной почты;
- забирает отсканированные документы из выделенной папки и делит их на комплекты;

После этого DCS передает комплекты документов утилите DrxUtil.

3. Утилита DrxUtil вызывает сервисы интеллектуальной обработки документов DIRECTUM Ario и передает им полученные комплекты документов.

4. Интеллектуальная обработка документов. Сервисы DIRECTUM Ario:

- делят комплект на отдельные документы;
- распознают текст документов;
- классифицируют документы по типам;
- извлекают факты из текста документов.

5. Из сервисов DIRECTUM Ario по каждому обработанному документу утилита DrxUtil получает информацию о типе документа и извлеченные факты. После этого утилита обращается с запросом к серверу приложений DirectumRX.

6. Занесение документов в DirectumRX. Сервер приложений по каждому обработанному документу получает из сервисов DIRECTUM Ario содержимое документа, а из утилиты DrxUtil информацию о типе и извлеченные факты. После этого сервер приложений:

- создает документы в DirectumRX и заполняет их карточки фактами, извлеченными из текста;
- формирует задание на проверку документов для ответственного сотрудника.

7. Верификация документов. В DirectumRX ответственный сотрудник получает задание на проверку документов. В рамках задания он проверяет, что документы занесены в полном объеме и с корректно заполненными карточками, после чего отправляет документы на дальнейшую обработку.

Обработка поступивших документов

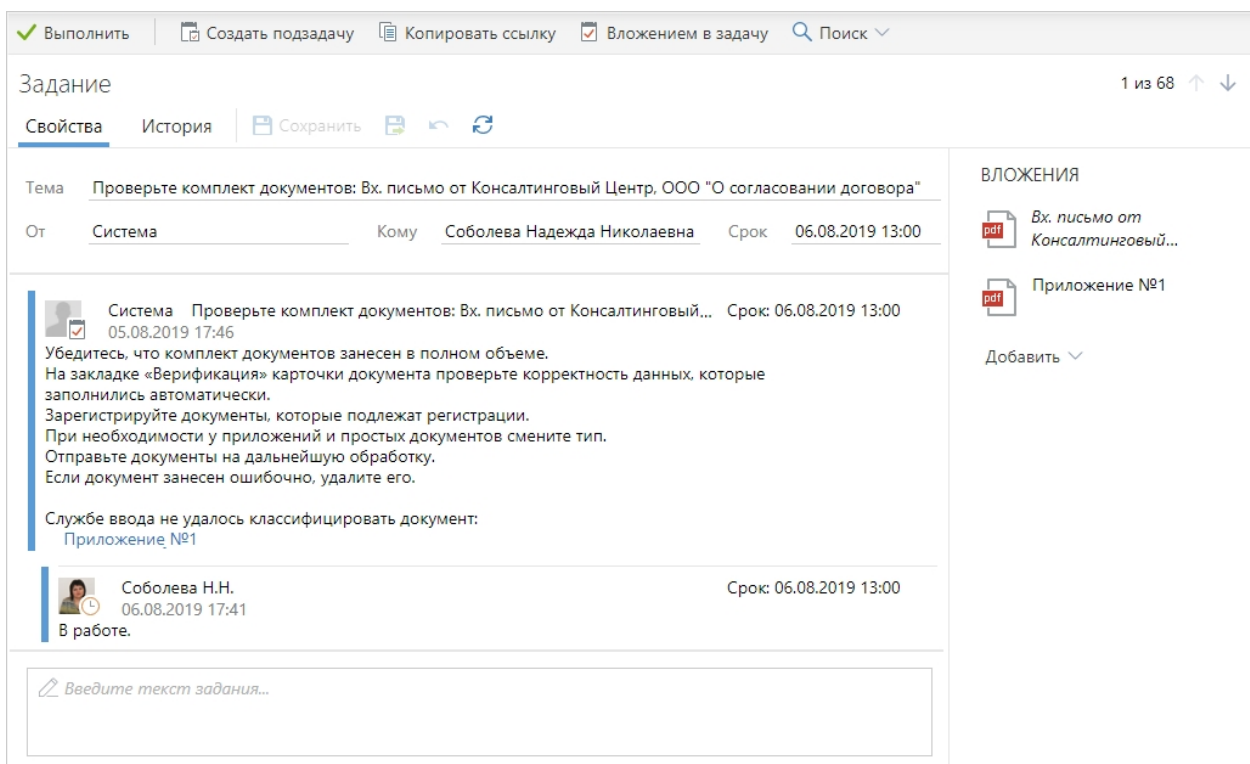
Для обеспечения корректного учета, поиска, согласования, рассмотрения и исполнения все документы после занесения в систему поступают на обработку ответственному сотруднику. Ответственный определяется согласно настройкам системы.

Совет. Перед обработкой документов ознакомьтесь с [особенностями занесения документов](#) в систему.

В качестве примера обработки документов рассмотрим работу секретаря с входящей корреспонденцией, которая поступает от контрагентов в бумажном виде.

При поступлении документов:

1. Подготовьте комплекты для сканирования. Отделите один комплект от другого белым листом.
2. Поместите комплекты документов в потоковый сканер. После сканирования система обработает документы и через некоторое время сформирует задания на проверку для каждого комплекта. Задания поступят в папку потока «На обработку».
3. Откройте задание на проверку и [убедитесь](#), что комплект занесен в систему в полном объеме.



4. [Верифицируйте](#) поступившие документы.
5. [Зарегистрируйте](#) документы, которые подлежат регистрации. Подробнее см. руководство пользователя DirectumRX, раздел «[Регистрация документов](#)». После этого в карточке документа группе полей «Жизненный цикл» в поле **Верификация** установится значение **Завершена**.
Если документ не подлежит регистрации, в карточке документа в поле **Верификация** вручную укажите значение **Завершена**. После завершения верификации и закрытия карточки документа закладка «Верификация» перестает отображаться и пропадает выделение полей цветом.
Если в комплекте есть нумеруемые документы (накладные, акты, счета-фактуры и универсальные передаточные документы), ознакомьтесь с [особенностями присвоения номера](#).
6. Отправьте документы на дальнейшую обработку, например, на рассмотрение или исполнение. Для этого в карточке документа на панели действий в выпадающем списке **Отправка** выберите подходящий пункт.

7. Выполните задание на проверку документов.

ПРИМЕЧАНИЕ. Если ранее верификация не была завершена (в карточке документа в поле **Верификация** осталось значение **В процессе**), то после выполнения задания и всех подзадач, которые из него отправлены, она завершится автоматически при выполнении фонового процесса «Документооборот. Завершение процесса верификации».

Проверка полноты комплекта

Если вы обнаружили, что документ добавлен в комплект по ошибке, удалите его. Подробнее см. руководство пользователя DirectumRX, раздел [«Удаление документов»](#).

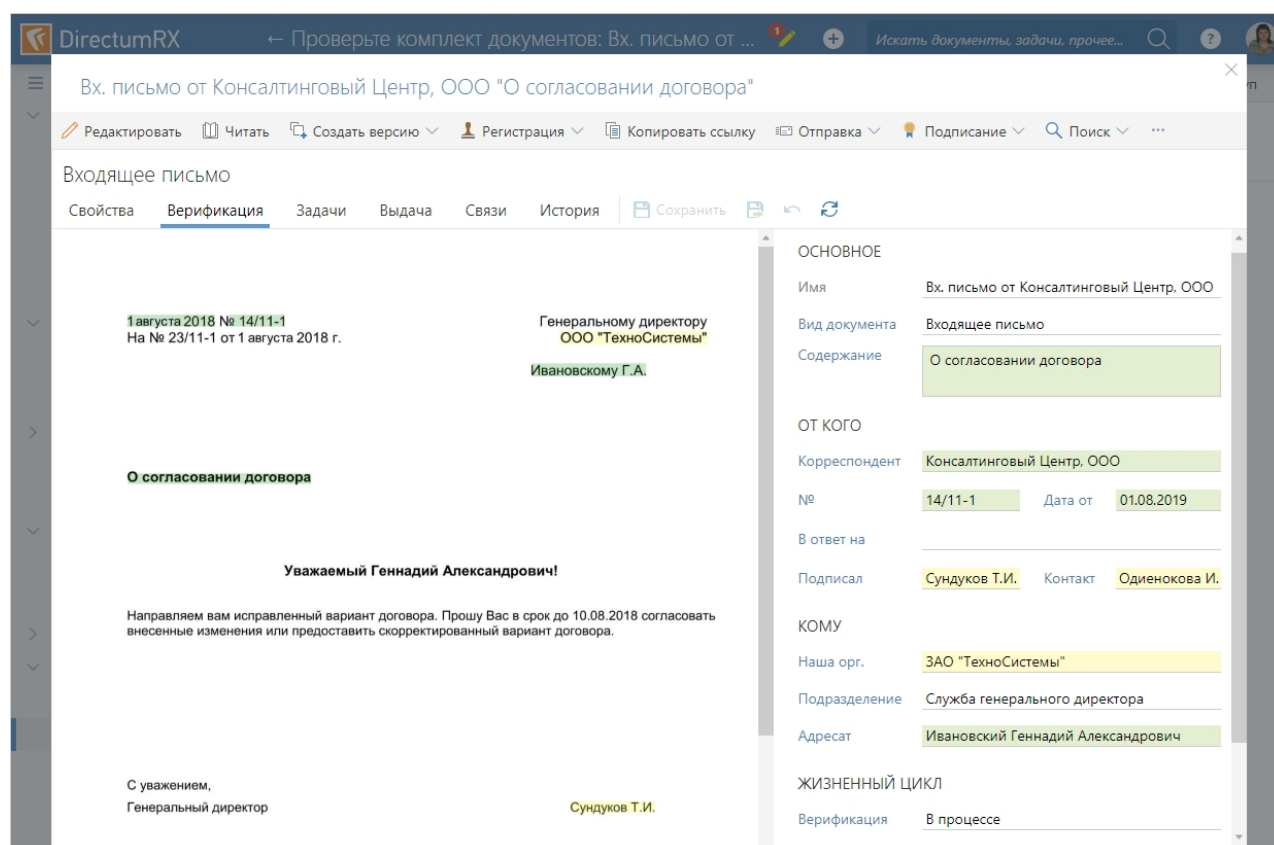
Если в комплекте не хватает документа, например, он ошибочно попал в другой комплект, то воспользуйтесь одним из способов:

- удалите все документы, вложенные в задание на проверку, и выполните его. В соответствующий комплект бумажных экземпляров добавьте необходимый документ. Заново отсканируйте и обработайте весь комплект;
- продолжите обработку этого комплекта, а затем отсканируйте необходимый документ отдельно. После получения задания на проверку, обработайте документ. Свяжите его как **Приложение** с ведущим документом нужного комплекта, либо свяжите с ним остальные документы комплекта, если он сам является ведущим.

Верификация документов

Для каждого документа из [задания на обработку](#):

1. Убедитесь, что содержимое занесено корректно и соответствуют бумажному экземпляру. Для этого откройте карточку документа. На открывшейся закладке «Верификация» посмотрите содержимое документа в окне предпросмотра:



Если документ занесен некорректно, удалите его. Подробнее см. руководство пользователя DirectumRX, раздел [«Удаление документов»](#).

Заново отсканируйте удаленный документ. После получения задания на проверку, обработайте документ и свяжите как **Приложение** с ведущим документом нужного комплекта, либо свяжите с ним остальные документы комплекта, если он сам является ведущим.

2. Проверьте в карточке корректность данных, которые заполнились автоматически. Для этого на закладке «Верификация» сравните текст в области предпросмотра и значения полей справа.

Факты, извлеченные из текста документа, выделены цветом в области предпросмотра и в полях карточки. **Зеленым** выделяются факты, которые с большой долей вероятности извлечены верно, а **желтым** – факты, на которые нужно обратить особое внимание.

Если в поле указано некорректное значение, исправьте его вручную.

ПРИМЕЧАНИЕ. При изменении значения поля, которое заполнено автоматически, выделение цветом пропадает.

3. Перейдите на закладку «Свойства» и заполните недостающие данные о документе. Например, укажите содержание, дело, в которое помещен документ, и дату помещения.

Если обрабатываете входящее письмо, проверьте, что в карточке верно указан способ доставки. Если документ занесен со сканера, то в поле **Способ доставки** автоматически указывается значение **Почта**, если по электронной почте – значение **Эл. почта**.

Если документ не подлежит регистрации, в поле **Верификация** укажите значение **Завершена**.

Особенности ввода документов в систему




Документы со сканера

Система распознает входящие письма, входящие счета, накладные, акты, счета-фактуры, универсальные передаточные документы и заносит их в систему с соответствующим типом.

Остальные документы и документы, у которых не удалось определить тип, заносятся как простые. Подробнее см. руководство пользователя DirectumRX, раздел [«Создание простого документа»](#). Для них указывается название:

- «Приложение № <порядковый номер в комплекте>», если у ведущего документа в комплекте удалось определить тип;
- «Простой документ № <порядковый номер в комплекте>», если не удалось определить тип ведущего документа в комплекте.

В тексте [задания на обработку поступивших документов](#) указывается, что службе ввода не удалось их классифицировать.

Тема Проверьте комплект документов: Вх. письмо от Консалтинговый Центр, ООО "О согласовании договора"			
От Система	Кому Соболева Надежда Николаевна	Срок 06.08.2019 13:00	
<div><div> Система Проверьте комплект документов: Вх. письмо от Консалтинговый... 05.08.2019 17:46</div><div>Убедитесь, что комплект документов занесен в полном объеме. На закладке «Верификация» карточки документа проверьте корректность данных, которые заполнились автоматически. Зарегистрируйте документы, которые подлежат регистрации. При необходимости у приложений и простых документов смените тип. Отправьте документы на дальнейшую обработку. Если документ занесен ошибочно, удалите его.</div></div>			
<div>Службе ввода не удалось классифицировать документ: Приложение №1</div>			
			ВЛОЖЕНИЯ <div><div> Вх. письмо от Консалтинговый...</div><div> Приложение №1</div><div>Добавить ▾</div></div>

Для корректного учета таких документов в системе смените их тип вручную. Подробнее см. руководство пользователя DirectumRX, раздел [«Смена типа простого документа»](#).

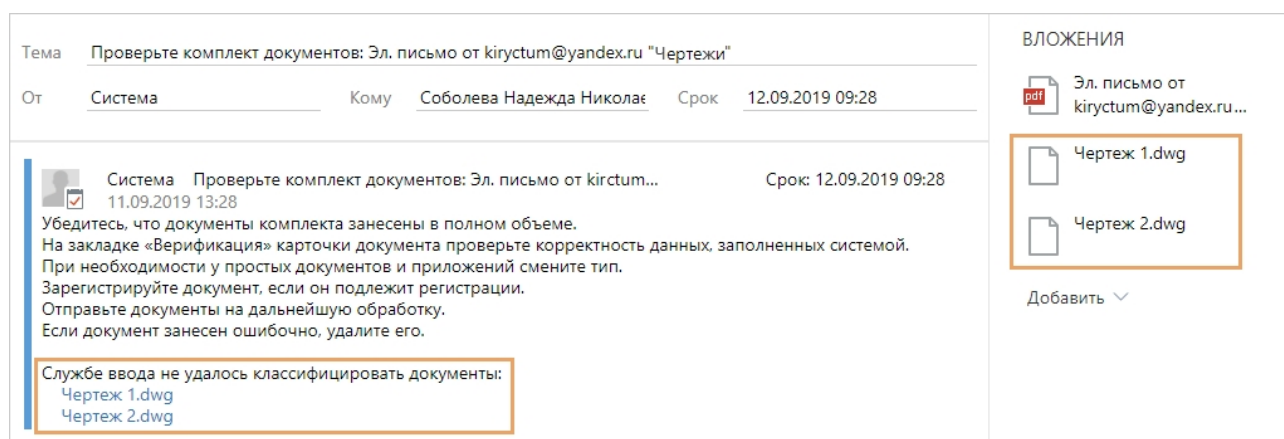
Например, если комплект состоит из сопроводительного письма, договора и технического задания, то письмо заносится как документ с типом «Входящее письмо», а договор и техническое задание как простые документы. Для договора смените тип вручную, а техническое задание оставьте простым документом, если для него в организации не предусмотрен отдельный тип документа.

Все отсканированные документы заносятся в систему в формате PDF.

Письма и документы с электронной почты

Письма и документы с электронной почты заносятся в систему аналогично [документам со сканера](#) с некоторыми особенностями:

- текст электронного письма преобразуется в формат PDF и заносится как простой документ. Если в тексте письма есть изображения, то они в документ не заносятся;
- документы, поступившие в формате JPG, JPEG, PNG, BMP, GIF, TIF, TIFF, DOC, DOCX, DOT, DOTX, RTF, ODT, OTT, TXT, XLS, XLSX, ODS PDF преобразуются в формат PDF. Документы других форматов заносятся в систему как простые документы в том формате, в котором поступили:






Возврат документов от контрагентов

В рамках согласования бумажный экземпляр документа может быть отправлен контрагенту на подписание. В этом случае нужно проконтролировать возврат в нашу организацию бумажного документа и занести в систему подписанный экземпляр. Подробнее см. руководство пользователя DirectumRX, раздел [«Контроль возврата документов от контрагентов»](#).

Документ, отправленный контрагенту, мог быть промаркирован штрихкодом. Подробнее см. руководство пользователя DirectumRX, раздел [«Печать штрихкода»](#). В этом случае при поступлении документа система находит его по штрихкоду и добавляет содержимое в новую версию. Данные в карточке повторно не заполняются.

В тексте задания указывается, что найден документ по штрихкоду:

Тема <u>Проверьте документ: Акт №1 от 05.09.2019 АБРАЗИВ, ООО</u>			ВЛОЖЕНИЯ		
От	Система	Кому	Соболева Надежда Николае	Срок	12.09.2019 10:33
<div> Система Проверьте документ: Акт №1 от 05.09.2019 АБРАЗИВ, ООО 11.09.2019 14:33</div> <p>Убедитесь, что документы комплекта занесены в полном объеме. На закладке «Верификация» карточки документа проверьте корректность данных, заполненных системой. При необходимости у простых документов и приложений смените тип. Зарегистрируйте документ, если он подлежит регистрации. Отправьте документы на дальнейшую обработку. Если документ занесен ошибочно, удалите его.</p> <div><p>В системе найден документ по штрихкоду. Создана новая версия. Акт №1 от 24.05.2019 АБРАЗИВ, ООО</p><p> Снигирев П.В.</p></div>			<div> Акт №1 от 05.09.2019 АБРАЗИВ, ООО</div> <p>Добавить ▾</p>		

Если от контрагента возвращен акт, накладная или УПД, то сотрудник, который указан ответственным в карточке документа, получит уведомление о поступлении документа.

Связи документов

Документы комплекта автоматически связываются между собой следующим образом:

1. Определяется ведущий документ:

- если в комплекте есть входящее письмо, то оно будет ведущим документом. Если писем несколько, то ведущий документ – первое письмо из комплекта;
- если входящего письма нет, то далее аналогично поочередно проверяется наличие в комплекте акта, накладной и счета-фактуры;
- если в комплекте нет входящего письма, акта, накладной и счета-фактуры, то ведущий документ – первый из занесенных в систему.

2. Проверяется тип ведущего документа:

- если у ведущего документа удалось распознать тип, то другие документы комплекта связываются с ним как **Приложение**. Документы, у которых не удалось распознать содержимое (занесены в неподдерживаемом формате или нечитаемые), связываются с ведущим как **Прочее**;
- если не удалось распознать тип ведущего документа и он занесен в систему как простой, то все документы комплекта связываются с ним как **Прочее**.

Пример 1. Комплект состоит из входящего письма, договора и технического задания. В этом случае ведущий документ – письмо, а договор и техническое задание связаны с ним как **Приложение**.

Пример 2. Комплект состоит из универсального передаточного документа (УПД) и договора. В этом случае, если в систему первым занесен:

- УПД, то он ведущий документ, а договор связан с ним как **Приложение**;
- договор, то он ведущий документ, а УПД связан с ним как **Прочее**.

Пример 3. Комплект получен по электронной почте и состоит из электронного письма, акта и файла в формате DWG. В этом случае ведущий документ – акт, электронное письмо (простой документ) связан с ним как **Приложение**, а документ в формате DWG связан как **Прочее**.

Присвоение номера документам

При поступлении нумеруемого документа (накладной, акта, счета-фактуры, универсального передаточного документа) возможны ситуации, когда:



- для данного вида документов есть подходящий журнал и настройки регистрации. Подробнее см. руководство пользователя DirectumRX, разделы [«Регистрация и нумерация»](#), [«Журналы регистрации»](#), [«Настройки регистрации»](#).

В этом случае система автоматически присвоит ему номер и дату, согласно заданным настройкам.

При этом, если номер документа определился с низкой вероятностью, то в карточке документа указывается значение **???**, а если с низкой вероятностью определилась дата, то вместо нее указывается значение **01.01.1753**. В этих случаях поля карточки остаются доступны для изменения. Измените номер и дату вручную.

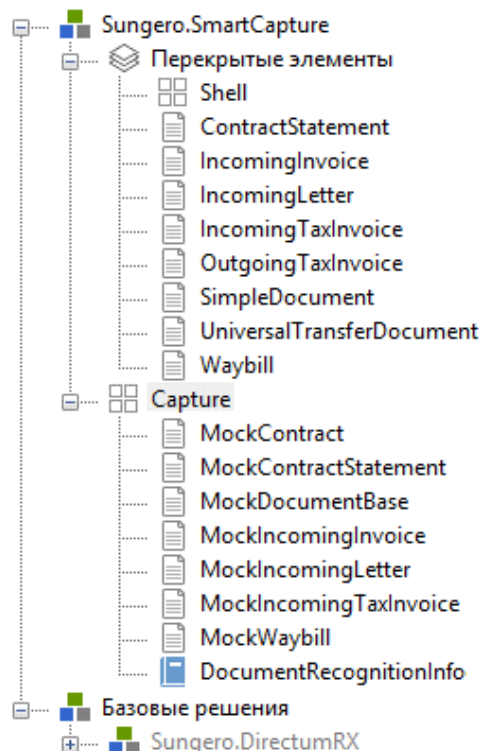
- системе не удалось подобрать подходящие настройки регистрации данного вида документов. Например, нет подходящего журнала и настройки регистрации, нужные записи закрыты, либо есть несколько подходящих журналов.

В тексте задания на обработку указывается для какого документа не удалось подобрать настройки. В этом случае обратитесь к администратору для корректной настройки системы.

Тема <u>Проверьте документ: С/ф полученный от Перевозофф, АО</u>			ВЛОЖЕНИЯ
От <u>Система</u>	Кому <u>Соболева Надежда Николае</u>	Срок <u>29.08.2019 16:30</u>	 С/ф полученный от Перевозофф, АО
<div><div><div>Система 29.08.2019 12:30</div></div><div>Проверьте документ: С/ф полученный от Перевозофф, АО Срок: 29.08.2019 16:30</div></div> <p>Убедитесь, что документы комплекта занесены в полном объеме. На закладке «Верификация» карточки документа проверьте корректность данных, заполненных системой. При необходимости у простых документов и приложений смените тип. Зарегистрируйте документ, если он подлежит регистрации. Отправьте документы на дальнейшую обработку. Если документ занесен ошибочно, удалите его.</p> <div>Для документа с видом "Счет-фактура полученный" не удалось подобрать настройку регистрации. Обратитесь к администратору системы. <u>С/ф полученный от Перевозофф, АО</u></div>			

Архитектура решения

В шаблоне «Интеллектуальная обработка документов» базовая логика интеллектуальной обработки документов реализована в решении Sungero.SmartCapture. Оно состоит из [модуля Capture](#) и [перекрытых элементов](#) стандартного решения DirectumRX.



Состав модуля Capture

Сторонняя библиотека ArioExtensions. Позволяет из прикладного кода DirectumRX вызывать методы сервисов DIRECTUM Ario для распознавания текста, классификации документов и извлечения фактов из текста. Описание методов API сервисов DIRECTUM Ario, примеры запросов и ответов доступны по [ссылке](#).

Справочник DocumentRecognitionInfo. Содержит факты, извлеченные сервисами DIRECTUM Ario из текста документов.

Документы для демонстрации возможностей решения:

- базовый демонстрационный документ (MockDocumentBase);
- договор (MockContract)
- акт выполненных работ (MockContractStatement);
- входящий счет на оплату (MockIncomingInvoice).
- входящее письмо (MockIncomingLetter);
- счет-фактура полученный (MockIncomingTaxInvoice);
- накладная (MockWayBil).

В отличие от других документов DirectumRX, все поля в карточках данных документов заполняются текстовыми значениями, а не ссылками на записи справочников. Например, поле **Контрагент** сразу заполняется значением, извлеченным из текста, и соответствующее значение в справочнике **Организации** уже не ищется. Это позволяет быстро продемонстрировать возможности шаблона технического решения «Интеллектуальная обработка документов» на реальных документах организации при работе в демо-режиме.

Функция SendToResponsible модуля Capture. Содержит логику формирования задачи на проверку документов.

Перекрытые элементы

Документы стандартной прикладной части DirectumRX:

- акт выполненных работ (ContractStatement);
- входящий счет на оплату (IncomingInvoice);
- входящее письмо (IncomingLetter);
- счет-фактура полученный (IncomingTaxInvoice);
- счет-фактура выставленный (OutgoingTaxInvoice);
- простой документ (SimpleDocument);
- универсальный передаточный документ (UniversalTransferDocument);
- накладная (WayBill).

Для перекрытий документов в карточку добавлена закладка «Верификация». На ней размещен контрол «Предпросмотр» и основные поля с информацией о документе. Также добавлено выделение цветом в тексте документа и в полях карточки тех фактов, которые сервисы DIRECTUM Ario извлекли из содержимого документа.

Модуль Shell. В перекрытии модуля в папке потока **На обработку** в событии **Получение данных** добавлена логика для помещения в папку потока заданий на проверку документов.

Установка и настройка решения

Для работы решения «Интеллектуальная обработка документов» необходимы:

- система DirectumRX 3.2 и выше;
- служба ввода документов (DCS);
- сервисы интеллектуальной обработки документов DIRECTUM Ario.

Установка решения состоит из этапов:

1. Установка решения и внешнего окружения в [контуре для разработки и тестирования](#).
2. [Адаптация решения](#) под бизнес-задачи организации.
3. Установка решения и внешнего окружения в [продуктивном контуре](#).

В дальнейшем решение может повторно дорабатываться в контуре для разработки и тестирования. В этом случае в продуктивной системе разработку необходимо [обновить](#).

Контур для разработки и тестирования

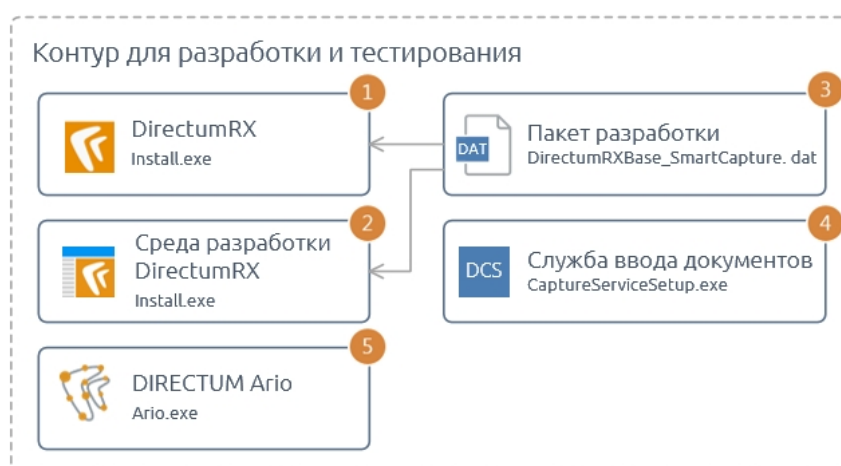
В контур для разработки и тестирования входит среда разработки и локальная система DirectumRX, установленные на компьютере разработчика. Используется для проверки работы решения, адаптации под задачи организации и тестирования внесенных изменений.

Установка решения и внешнего окружения

Для установки необходимо:

Компонент	Расположение	Имя файла
Программа установки DirectumRX	Комплект поставки DirectumRX	Install.exe
Файл пакета разработки «Интеллектуальный захват документов»	Комплект поставки решения «Интеллектуальная обработка документов»	DirectumRXBase_SmartCapture.dat
Программа установки сервисов DIRECTUM Ario	Комплект поставки сервисов DIRECTUM Ario	Ario.exe
Программа установки DCS	Комплект поставки DCS	CaptureServiceSetup.exe

Порядок установки:



1. Установите систему DirectumRX. Подробнее см. инструкцию по установке, входит в комплект поставки DirectumRX.
2. Установите среду разработки DirectumRX. Подробнее см. инструкцию по установке, входит в комплект поставки DirectumRX.
3. Опубликуйте файл пакета разработки DirectumRXBase_SmartCapture.dat:
 - в систему DirectumRX. Подробнее см. в справке DirectumRX, раздел [«Утилита DeploymentTool»](#).
 - в среду разработки. Подробнее см. в руководстве по среде разработки, раздел [«Импорт разработки»](#).

Пакет разработки входит в комплект поставки решения «Интеллектуальная обработка документов».
4. Установите службу ввода документов (DCS). Подробнее см. инструкцию по установке, раздел «Установка служб ввода», входит в комплект поставки DCS.
5. Установите сервисы DIRECTUM Ario. Подробнее см. инструкцию по установке, входит в комплект поставки сервисов DIRECTUM Ario.

Настройка решения и внешнего окружения

При настройке используются следующие компоненты:

Компонент	Путь до файла по умолчанию
DCS	%PROGRAMFILES%\DIRECTUM Company\DIRECTUM Capture Service
Утилита DrxUtil	%PROGRAMFILES%\DIRECTUM Company\DirectumRX\Utilities\DrxUtil\DrxUtil.exe
Модель классификации по первым страницам	Входит в комплект поставки решения «Интеллектуальная обработка документов», файл model\first_page_classifier_model.zip
Модель классификации по типам документов	Входит в комплект поставки решения «Интеллектуальная обработка документов», файл model\kinds_klassifier_model.zip
Bat-файл с настройками отправки документов в DirectumRX	Входит в комплект поставки решения «Интеллектуальная обработка документов», файл conf\SendToRX.bat.example
Конфигурационный файл настроек DCS	%PROGRAMFILES%\DIRECTUM Company\DIRECTUM Capture Service\bin\CaptureService.exe.config
Пример конфигурационного файла настроек DCS для отправки документов в DirectumRX	Входит в комплект поставки решения «Интеллектуальная обработка документов», файл conf\CaptureService.exe.config.example

Порядок настройки:

1. [Создайте](#) в сервисах DIRECTUM Ario [классификаторы](#) по первым страницам и по типам документов.
2. [Импортируйте](#) в классификаторы модели классификации по первым страницам и по типам документов, которые входят в комплект поставки решения «Интеллектуальная обработка документов».
3. Определите:
 - папки, из которых будут захватываться отсканированные бумажные документы;
 - ящики электронной почты, из которых будут захватываться документы и письма.
4. [Настройте](#) службу ввода документов.

5. [Задайте](#) настройки DirectumRX для работы с сервисами DIRECTUM Ario.
6. [Проверьте](#) захват документов из папок и с электронной почты.

Классификатор и модели классификации

Классификация – этап обработки документа, при котором сервисы DIRECTUM Ario по тексту определяют, к какому классу относится документ. В качестве классов используются, например, типы документов: входящее письмо, входящий счет, накладная, акт, счет-фактура, универсальный передаточный документ.

В результате классификации формируется модель классификации – совокупность файлов, которые содержат набор признаков, характерных для каждого класса, а также скрипты и алгоритмы, выполняющие классификацию.

В комплект поставки решения «Интеллектуальная обработка документов» входит:

- модель классификации по первым страницам (first_page_classifier_model.zip). С ее помощью определяется, является ли страница файла первой в документе. Это позволяет разделить файл на несколько документов. Модель поставляется в виде ZIP-архива;
- модель классификации по типам документов (kinds_classifier_model.zip). С ее помощью по тексту определяется, к какому типу относится документ. Модель поставляется в виде ZIP-архива.

ВАЖНО. Готовая модель классификации по типам документов подойдет для первоначальной настройки и тестирования решения. В дальнейшем для более точной классификации необходимо [обучить модель классификатор](#) на реальных документах организации.

Чтобы сервисы DIRECTUM Ario делили комплекты на отдельные документы и определяли их типы, [создайте](#) в сервисах:

- классификатор с моделью классификации по первым страницам;
- классификатор с моделью классификации по типам документов.



Проверка захвата документов

Чтобы проверить:

- захват отсканированных бумажных документов, поместите файлы в папку, для которой настроен захват службой DCS. Путь до папки указан в конфигурационном файле CaptureService.exe.config в секции **FileSystem** в параметре **FullPath**. После того, как DCS захватит документы и начнет обработку, они исчезнут из папки;
- захват электронных документов, отправьте документы на адрес электронной почты, для которого настроен захват службой DCS. Адрес указан в конфигурационном файле CaptureService.exe.config в секции **Mail** в параметре **User**. После того, как DCS захватит документы и начнет обработку, они исчезнут из папки «Входящие» почтового ящика.

Если захват прошел успешно, то в DirectumRX через некоторое время сформируется задание на проверку комплекта документов. Задание получит сотрудник, указанный ответственным за проверку комплекта в исполняемом файле SendToRX.bat, который используется при [настройке DCS](#). Убедитесь, что ответственному сотруднику пришло задание на проверку.

Если задание не сформировалось, посмотрите ошибки в лог-файлах DCS и утилиты DrxUtil.

Примечание. При успешном захвате документов лог-файлы DCS не заполняются.

Адаптация решения под бизнес-задачи организации

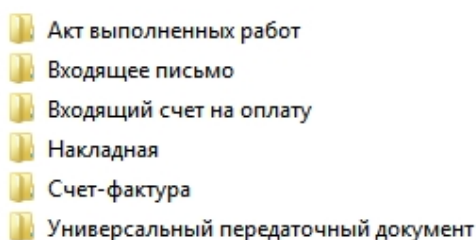
Решение «Интеллектуальная обработка документов» имеет статус шаблона и реализует пример интеллектуальной обработки нескольких типов документов. Модель классификации по типам документов, которая входит в комплект поставки, также представляет собой стартовый пример работы решения. Этого достаточно для первоначальной проверки и настройки решения в контуре для разработки и тестирования.

Чтобы использовать решение в продуктивном контуре, повысить точность классификации и извлечения фактов, необходимо адаптировать решение под задачи организации и обучить классификатор на реальных документах организации. Для этого:



1. В среде разработки DirectumRX адаптируйте решение под необходимые бизнес-задачи. Например, добавьте интеллектуальную обработку для новых типов документов, доработайте задачу на обработку документов, измените логику заполнения полей карточке извлеченными фактами и т.д. Подробнее см. [пример](#) добавления интеллектуальной обработки для авансового отчета.
2. Подготовьте документы для обучения. Потребуется минимум 100 документов каждого типа. Чем больше документов, тем выше точность классификации и дольше длится обучение.

На сервере, где установлены сервисы DIRECTUM Ario, в сетевой папке создайте структуру папок документов каждого типа.



Требования к размещению документов для обучения:

- путь до папок с документами для обучения должен состоять из латинских символов без пробелов, использование кириллицы не допускается;
 - названия папок должны быть содержательными и написаны кириллицей. Типы документов, которые создаются при обучении, получают названия соответствующих папок. Например, если папка называется «Акт выполненных работ», то соответствующий тип документов будет назван «Акт выполненных работ»;
 - размещайте в папках только документы соответствующего типа. Например, в папке «Входящее письмо» должны лежать только входящие письма. Если в папке содержатся документы других типов, точность классификации документов и распознавания фактов снижается.
3. [Обучите классификатор](#) по типам документов на подготовленных документах.
 4. [Проверьте захват документов](#).

Если точность классификации документов и распознавания фактов недостаточно высокая:

1. Проверьте, что в каждой папке с документами для обучения находятся только документы нужного типа. Например, в папке «Входящее письмо» должны находиться только входящие письма, в папке «Накладная» только накладные и т.д.
2. Увеличьте количество документов каждого типа.
3. Повторите обучение классификатора.
4. Повторно проверьте захват документов.

Если при одном из прошлых обучений точность классификации документов и распознавания фактов была выше, чем при повторном обучении, можно откатить изменения и [опубликовать](#) в классификатор модель классификации с самыми высокими показателями. В результате данная модель будет использоваться для обработки документов.

Когда захват документов прошел успешно, а точность классификации и распознавания фактов вас устраивает, подготовьте **комплект для переноса разработки в продуктивный контур**. Для этого:

- экспортируйте пакет разработки с исходными кодами из среды разработки DirectumRX. Подробнее см. руководство по среде разработки DirectumRX, раздел [«Экспорт разработки»](#);
- [экспортируйте](#) обученную модель классификации из сервисов DIRECTUM Ario.

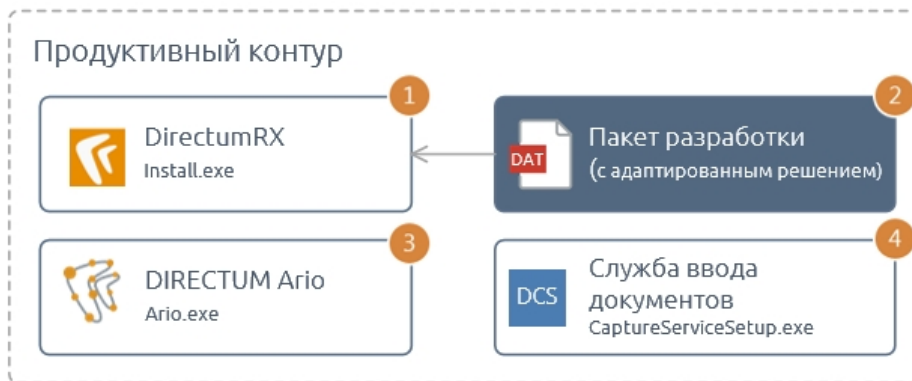
Продуктивный контур

Установка решения и внешнего окружения

Для установки необходимо:

Компонент	Расположение	Имя файла
Программа установки DirectumRX	Комплект поставки DirectumRX	Install.exe
Файл пакета разработки	Комплект для переноса решения в продуктивную систему	Имя указывается при экспорте пакета разработки из среды разработки
Программа установки сервисов DIRECTUM Ario	Комплект поставки сервисов DIRECTUM Ario	Ario.exe
Программа установки DCS	Комплект поставки DCS	CaptureServiceSetup.exe

Порядок установки:



1. Установите систему DirectumRX. Подробнее см. инструкцию по установке, входит в комплект поставки DirectumRX.
2. Опубликуйте в систему DirectumRX файл пакета разработки из комплекта для переноса решения в продуктивную систему. Подробнее о публикации см. в справке DirectumRX, раздел [«Утилита DeploymentTool»](#).
3. Установите сервисы DIRECTUM Ario. Подробнее см. инструкцию по установке, входит в комплект поставки сервисов DIRECTUM Ario.
4. Установите службу ввода документов (DCS). Подробнее см. инструкцию по установке, раздел «Установка служб ввода», входит в комплект поставки DCS.

Настройка решения и внешнего окружения

Порядок настройки в продуктивном контуре аналогичен настройке в [контуре для разработки и тестирования](#) с некоторыми отличиями:

- при [создании](#) классификатора по типам документов в качестве модели классификации используйте обученную модель из комплекта для переноса решения в продуктивную систему;
- при настройке папок, из которых будут захватываться отсканированные бумажные документы, настройте сканеры на помещение отсканированных документов в данные папки.

Обновление продуктивной системы

Если после первоначальной настройки решение повторно [дорабатывалось](#), то обновите разработку в продуктивном контуре. Для этого:

1. Опубликуйте в систему DirectumRX файл пакета разработки с доработанным решением. Подробнее см. в справке DirectumRX, раздел [«Утилита DeploymentTool»](#).
2. [Импортируйте](#) в сервисы DIRECTUM Ario обученную модель классификации по типам документов.
3. [Проверьте](#) захват документов из папки и с электронной почты.

Приложение 1. Команды командной строки для настройки решения

Настройка DirectumRX для работы с DIRECTUM Ario

Команда используется для подключения системы DirectumRX к сервисами DIRECTUM Ario. Выполняется при настройке решения и внешнего окружения в [контуре для разработки и тестирования](#) и в [продуктивном контуре](#).

Синтаксис

```
[ПУТЬ_ДО_DRXUTIL] -n <login> -p <password> -f Sungero.Capture.Module.SetCaptureMainSettings  
"<ARIO_URL>" <MIN_FACT_PROBABILITY> <TRUSTED_FACT_PROBABILITY>
```

[ПУТЬ_ДО_DRXUTIL] – путь до утилиты DrxUtil.

-n <login> – пользователь, от имени которого запускается утилита.

-p <password> – пароль пользователя, от имени которого запускается утилита.

-f Sungero.Capture.Module.SetCaptureMainSettings – ключ и функция для настройки взаимодействия DirectumRX и сервисов DIRECTUM Ario.

"<ARIO_URL>" – адрес сервисов DIRECTUM Ario, который был указан при их установке.

<MIN_FACT_PROBABILITY> – минимальная вероятность для факта. Факты с вероятностью меньше указанной игнорируются при заполнении карточек документов. Например, 40.

<TRUSTED_FACT_PROBABILITY> Доверительная вероятность для факта. Например, 80. Влияет на подсветку свойств в режиме верификации:

- факты, определенные с вероятностью ниже указанного числа, подсвечиваются желтым цветом. При верификации документов сотруднику нужно обратить особое внимание на такие факты и перепроверить их;
- факты, определенные с вероятностью выше указанного числа, подсвечиваются зеленым цветом. При верификации документов сотруднику не обязательно перепроверять их.

Пример

```
D:\Projects\DirectumRX\bin\Debug\DrxUtil\DrxUtil.exe -n "Service user" -p 11111 -f  
Sungero.Capture.Module.SetCaptureMainSettings "http://smart:61100" 40 80
```

Создание классификатора

Команда используется для создания [классификатора](#) в сервисах DIRECTUM Ario. Выполняется при настройке системы и внешнего окружения в [контуре для разработки и тестирования](#) и в [продуктивном контуре](#).

Синтаксис

```
[ПУТЬ_ДО_DRXUTIL] -n <login> -p <password> -f Sungero.Capture.Module.CreateClassifier  
"<ИМЯ_КЛАССИФИКАТОРА>" "<МИНИМАЛЬНАЯ_ВЕРОЯТНОСТЬ>"
```

[ПУТЬ_ДО_DRXUTIL] – путь до утилиты DrxUtil.

-n <login> – пользователь, от имени которого запускается утилита.

-p <password> – пароль пользователя, от имени которого запускается утилита.

-f Sungero.Capture.Module.CreateClassifier – ключ и функция для создания классификатора сервисов DIRECTUM Ario.

<ИМЯ_КЛАССИФИКАТОРА> – имя создаваемого классификатора;

<МИНИМАЛЬНАЯ_ВЕРОЯТНОСТЬ> – минимальная вероятность соответствия документа одному из выделенных классов. Если классификатор определил, что вероятность соответствия документа каждому классу меньше указанного числа, то документ не будет отнесен ни к одному из классов. Например, если документ соответствует всем типам с вероятностью меньше указанного числа, то считается, что его тип не определен и он заносится в систему как простой документ. Возможное значение задается в интервале от 0 до 1. Рекомендуемое значение 0,75.

Примеры

Создание классификатора по первым страницам:

```
D:\Projects\DirectumRX\bin\Debug\DrxUtil\DrxUtil.exe -n Administrator -p 11111 -f  
Sungero.Capture.Module.CreateClassifier "Классификатор первых страниц" "0.75"
```

Создание классификатора по типам документов:

```
D:\Projects\DirectumRX\bin\Debug\DrxUtil\DrxUtil.exe -n Administrator -p 11111 -f  
Sungero.Capture.Module.CreateClassifier "Классификатор по типам документов" "0.75"
```

Импорт модели классификации

Команда используется для:

- загрузки модели классификации в новый классификатор. Выполняется при настройке решения и внешнего окружения в [контуре для разработки и тестирования](#) и в [продуктивном контуре](#);
- обновления модели классификации в существующем классификаторе. Выполняется при [обновлении продуктивной системы](#).

Синтаксис

```
[ПУТЬ_ДО_DRXUTIL] -n <login> -p <password> -f Sungero.Capture.Module.ImportClassifierModel  
"<ИМЯ_КЛАССИФИКАТОРА>" "<ПУТЬ_ДО_ФАЙЛА_МОДЕЛИ>"
```

[ПУТЬ_ДО_DRXUTIL] – путь до утилиты DrxUtil.

-n <login> – пользователь, от имени которого запускается утилита.

-p <password> – пароль пользователя, от имени которого запускается утилита.

-f Sungero.Capture.Module.ImportClassifierModel – ключ и функция для импорта модели классификации в классификатор сервисов DIRECTUM Ario.

<ИМЯ_КЛАССИФИКАТОРА> – имя классификатора, в который нужно импортировать модель;

<ПУТЬ_ДО_ФАЙЛА_МОДЕЛИ> – путь до zip-архива с моделью классификации.

Примеры

Импорт модели классификации по первым страницам:

```
D:\Projects\DirectumRX\bin\Debug\DrxUtil\DrxUtil.exe -n Administrator -p 11111 -f  
Sungero.Capture.Module.ImportClassifierModel "Классификатор первых страниц"  
"C:\Models\Import\first_page_classifier_model.zip"
```

Импорт модели классификации по типам документов:

```
D:\Projects\DirectumRX\bin\Debug\DrxUtil\DrxUtil.exe -n Administrator -p 11111 -f  
Sungero.Capture.Module.ImportClassifierModel "Классификатор по типам документов"  
"C:\Models\Import\first_page_classifier_model.zip"
```

Обучение классификатора

Команда используется для обучения классификатора на реальных документах организации, чтобы «научить» его определять новые классы и повысить точность классификации. Выполняется на этапе [адаптации](#) решения под бизнес-задачи организации.

Перед обучением подготовьте документы и разместите их в сетевой папке на сервере, где установлены сервисы DIRECTUM Ario, в соответствии с [требованиями](#) к размещению документов для обучения.

Обучение можно выполнять несколько раз, чтобы повысить качество классификации документов. В результате каждого этапа обучения в классификаторе создается новая модель.

Синтаксис

```
[ПУТЬ_ДО_DRXUTIL] -n <login> -p <password> -f Sungero.Capture.Module.TrainClassifierModel  
<ИМЯ_КЛАССИФИКАТОРА> <СЕТЕВОЙ_ПУТЬ_ДО_ПАПКИ>
```

[ПУТЬ_ДО_DRXUTIL] – путь до утилиты DrxUtil.

-n <login> – пользователь, от имени которого запускается утилита.

-p <password> – пароль пользователя, от имени которого запускается утилита.

-f Sungero.Capture.Module.TrainClassifierModel – ключ и функция для обучения классификатора.

<ИМЯ_КЛАССИФИКАТОРА> – имя классификатора, из которого нужно экспортировать модель.

<СЕТЕВОЙ_ПУТЬ_ДО_ПАПКИ> – путь до папки с документами для обучения.

Пример

```
D:\Projects\DirectumRX\bin\Debug\DrxUtil\DrxUtil.exe -n Administrator -p 11111 -f
Sungero.Capture.Module.TrainClassifierModel "Классификатор первых страниц"
"\\server\DataSets\DatasetArio"
```

Просмотр моделей, используемых в классификаторе

Команда используется для отображения списка моделей классификатора с указанием их идентификаторов и текущей (активной) модели.

Синтаксис

```
[ПУТЬ_ДО_DRXUTIL] -n <login> -p <password> -f Sungero.Capture.Module.ShowClassifierModels
"<ИМЯ_КЛАССИФИКАТОРА>"
```

[ПУТЬ_ДО_DRXUTIL] – путь до утилиты DrxUtil.

-n <login> – пользователь, от имени которого запускается утилита.

-p <password> – пароль пользователя, от имени которого запускается утилита.

-f Sungero.Capture.Module.ShowClassifierModels – ключ и функция для показа моделей, используемых в классификаторе.

<ИМЯ_КЛАССИФИКАТОРА> – имя классификатора, модели которого нужно посмотреть.

Пример

```
D:\Projects\DirectumRX\bin\Debug\DrxUtil\DrxUtil.exe -n Administrator -p 11111 -f
Sungero.Capture.Module.ShowClassifierModels "Классификатор первых страниц"
```

Результат выполнения:

```
Classifier "TypeClassifier" with Id 2151, created 06.09.2019 15:29:51, min probability 0,8. Models:
*CURRENT* Model with Id 2141, created 12.09.2019 13:39:33. Train set count 66, accuracy 0,875.
----- Model with Id 2140, created 12.09.2019 13:31:33. Train set count 66, accuracy 0,375.
----- Model with Id 2123, created 10.09.2019 14:16:12. Train set count 66, accuracy 0,875.
----- Model with Id 2122, created 10.09.2019 14:07:42. Train set count 66, accuracy 0,5.
----- Model with Id 2121, created 10.09.2019 13:55:42. Train set count 66, accuracy 1.
----- Model with Id 2120, created 10.09.2019 13:53:42. Train set count 66, accuracy 0,75.
----- Model with Id 2119, created 10.09.2019 13:46:42. Train set count 66, accuracy 0,5.
----- Model with Id 2118, created 10.09.2019 13:45:12. Train set count 66, accuracy 0,75.
----- Model with Id 2117, created 10.09.2019 13:45:12. Train set count 66, accuracy 0,75.
----- Model with Id 2116, created 10.09.2019 13:42:12. Train set count 66, accuracy 0,625.
----- Model with Id 2115, created 10.09.2019 13:39:12. Train set count 66, accuracy 1.
----- Model with Id 2114, created 10.09.2019 13:33:12. Train set count 66, accuracy 0,75.
```


Публикация модели классификации

Команда используется для назначения активной модели из числа моделей, загруженных в классификатор. Активная модель используется при обработке документов.

Выполняется на этапе [адаптации](#) решения под бизнес-задачи организации.

Перед публикацией [посмотрите список моделей](#) классификатора, чтобы узнать ID нужной модели.

Синтаксис

```
[ПУТЬ_ДО_DRXUTIL] -n <login> -p <password> -f Sungero.Capture.Module.PublishClassifierModel  
" <ИМЯ_КЛАССИФИКАТОРА> " " <ID_МОДЕЛИ> "
```

[ПУТЬ_ДО_DRXUTIL] – путь до утилиты DrxUtil.

-n <login> – пользователь, от имени которого запускается утилита.

-p <password> – пароль пользователя, от имени которого запускается утилита.

-f Sungero.Capture.Module.PublishClassifierModel – ключ и функция для публикации модели классификации.

<ИМЯ_КЛАССИФИКАТОРА> – имя классификатора.

<ID_МОДЕЛИ> – идентификатор модели классификации, которую нужно опубликовать.

Пример

```
D:\Projects\DirectumRX\bin\Debug\DrxUtil\DrxUtil.exe -n Administrator -p 11111 -f  
Sungero.Capture.Module.PublishClassifierModel "Классификатор первых страниц" "2111"
```

Экспорт модели классификации

Команда используется для экспорта из сервисов DIRECTUM Ario модели классификации, обученной на документах организации. Выполняется на этапе [адаптации](#) решения под бизнес-задачи организации при подготовке комплекта для переноса разработки в продуктивный контур.

Экспортированная обученная модель используется при [настройке](#) и [обновлении продуктивной системы](#).

Перед экспортом [посмотрите список моделей](#) классификатора, чтобы узнать ID нужной модели.

Синтаксис

```
[ПУТЬ_ДО_DRXUTIL] -n <login> -p <password> -f Sungero.Capture.Module.ExportClassifierModel  
<ИМЯ_КЛАССИФИКАТОРА> <ID_МОДЕЛИ> <ПУТЬ_ДЛЯ_СОХРАНЕНИЯ>
```

[ПУТЬ_ДО_DRXUTIL] – путь до утилиты DrxUtil.

-n <login> – пользователь, от имени которого запускается утилита.

-p <password> – пароль пользователя, от имени которого запускается утилита.

-f Sungero.Capture.Module.ExportClassifierModel – ключ и функция для экспорта модели классификации.

<ИМЯ_КЛАССИФИКАТОРА> – имя классификатора, из которого нужно экспортировать модель.

<ID_МОДЕЛИ> – идентификатор обученной модели классификации.

<ПУТЬ_ДЛЯ_СОХРАНЕНИЯ> – путь до папки, в которую будет экспортирован zip-архив с обученной моделью классификации.

Пример

```
D:\Projects\DirectumRX\bin\Debug\DrxUtil\DrxUtil.exe -n Administrator -p 11111 -  
fSungero.Capture.Module.ExportClassifierModel "Классификатор первых страниц" "2110"  
"C:\Models\Export\first_page_classifier_model zip"
```

Приложение 2. Настройка службы ввода документов

Примечание. Служба ввода документов по умолчанию устанавливается в папку %PROGRAMFILES%\DIRECTUM Company\DIRECTUM Capture Service, далее для нее используется сокращенное обозначение {DCS_DIR}.

Настройки службы ввода документов (DCS) хранятся в конфигурационном файле CaptureService.exe.config. Он находится в папке {DCS_DIR}\bin и заполнен значениями по умолчанию.

В комплекте поставки решения «Интеллектуальная обработка документов» есть конфигурационный файл CaptureService.exe.config.example. В нем задано большинство настроек, необходимых для захвата и отправки документов в DirectumRX. Он находится в папке conf и заполнен значениями по умолчанию.

Чтобы настроить захват документов DCS и отправку их в DirectumRX:

1. Скопируйте файл CaptureService.exe.config.example из папки conf комплекта поставки решения «Интеллектуальная обработка документов» в папку {DCS_DIR}\bin.
2. Переименуйте файл в CaptureService.exe.config и откройте его на редактирование.
3. Задайте настройки для [захвата документов из папки](#) и [с ящика электронной почты](#).
4. [Настройте пути до папок](#), в которых будут размещаться лог-файлы и временные файлы DCS.
5. Скопируйте файл SendToRX.bat.example из папки conf комплекта поставки решения «Интеллектуальная обработка документов» в папку {DCS_DIR}.
6. Переименуйте файл в SendToRX.bat и откройте его на редактирование.
7. В файле SendToRX.bat укажите:

```
{DRXUTIL_PATH} -n "{NAME}" -p {PASSWORD} -f Sungero.Capture.Module.ProcessCapturedPackage %1 %2 %3 %4 %5 {ID} "{CLASSIFIER}" "{TYPE_CLASSIFIER}"
```

{DRXUTIL_PATH} – путь до утилиты DrxUtil.

-n "{NAME}" – пользователь, от имени которого запускается утилита.

-p {PASSWORD} – пароль пользователя, от имени которого запускается утилита.

-f Sungero.Capture.Module.ProcessCapturedPackage – ключ и функция для настройки DCS.

%1 %2 %3 %4 %5 – параметры настроек службы DCS, оставьте их без изменения.

{ID} – ИД пользователя DirectumRX, которому будут поступать задания на обработку комплектов документов. Например, 81.

"{CLASSIFIER}" – имя модели классификации первых страниц. Например, first_page_classifier_model.

"{TYPE_CLASSIFIER}" – имя модели классификации по типам документов. Например, kinds_klassifier_model.

Пример:

```
"D:\Projects\DirectumRX\bin\Debug\DrxUtil\DrxUtil.exe" -n "Service user" -p 11111 -f Sungero.Capture.Module.ProcessCapturedPackage %1 %2 %3 %4 %5 81 "first_page_classifier_model" "kinds_klassifier_model"
```

COBET. Чтобы документы, занесенные со сканера, обрабатывал один сотрудник, а документы с электронной почты – другой, создайте два файла SendToRX.bat. В первом файле в параметре {ID} укажите ИД пользователя DirectumRX, который будет обрабатывать документы со сканера. Второй файл назовите, например, SendToRX2.bat и в параметре {ID} укажите ИД пользователя DirectumRX, который будет обрабатывать документы с почты.

8. [Настройте пути до папок](#), в которые будут помещаться временные файлы DirectumRX.
9. Запустите DCS из командной строки с помощью команды:

```
net start DirectumCaptureService
```

Настройка ввода документов из файловой системы

Чтобы настроить занесение документов в систему из выделенной папки, в файле CaptureService.exe.config в секции **FileSystem** в параметре **FullPath** укажите путь до нужной папки. По умолчанию указана папка {DCS_DIR}\FileCapture\Incoming.

ВАЖНО. Для работы службы ввода документов в параметрах конфигурационного файла фрагмент пути «{DCS_DIR}» в значениях по умолчанию нужно заменить на реальный путь до нужной папки.

Пример:

```
<FileSystem
  CaptureLegacyFiles="true"
  SenderLine="StreamLineToSystem">
  <clear/>
  <add FullPath="D:\DCS\FileCapture\Incoming"
    ...
</FileSystem>
```

Для детальной настройки способа занесение документов в систему из папки измените значения остальных параметров секции. Подробнее описание параметров см. в инструкции по установке DCS, раздел «Настройка модуля ввода из файловой системы». Входит в комплект поставки DCS.

Настройка ввода документов с электронной почты

Чтобы настроить занесение документов в систему из выделенного ящика электронной почты, в файле CaptureService.exe.config в секции **Mail** заполните параметры:

- **Server** – адрес почтового сервера;
- **User** – имя пользователя для подключения к почтовому серверу;
- **Password** – пароль пользователя для подключения к почтовому серверу.

Пример:

```
<Mail SenderLine="MailCaptureLine" >
  <clear />
  <add
    ...>
  <Server>
    <clear />
    <add Name="Type" Value="POP3"/>
    <add Name="Server" Value="mail.company.ru"/>
    <add Name="User" Value="dcsmail@company.ru"/>
    <add Name="Password" Value="password"/>
```

```
...
</Server>
<add>
</Mail>
```

Для детальной настройки способа занесение документов в систему с ящика электронной почты измените значения остальных параметров секции. Подробнее описание параметров см. в инструкции по установке DCS, раздел «Настройка модуля ввода с электронной почты». Входит в комплект поставки DCS.

Папки для лог-файлов и временных файлов службы DCS

Чтобы настроить:

- папку для хранения лог-файлов DCS, в файле CaptureService.exe.config в секции **appSettings** в параметре **key** укажите путь до нужной папки. По умолчанию указана папка {DCS_PATH}\Logs.

Пример:

```
<appSettings>
  <add key="logsFolder" value="D:\DCS\Logs"/>
</appSettings>
```

- папки для хранения временных файлов линий обработки документов, захваченных из выделенной папки и с ящика электронной почты, в файле CaptureService.exe.config в секции **Lines** для каждой линии в параметрах **WorkingFolder** укажите пути до нужных папок. По умолчанию указаны папки {DCS_PATH}\FileCapture\FileCaptureLineTemp и {DCS_PATH}\MailCapture\MailCaptureLineTemp.

Пример:

```
<Lines>
  <clear/>
  <add
    <!-- Линия для захвата документов из файловой системы. -->
    Name="FileCaptureLine"
    Packer="FileCapturePacker"
    Sender="AnyProgramSender"
    WorkingFolder="C:\DCS\FileCapture\FileCaptureLineTemp"/>
  <add
    <!-- Линия для захвата документов с ящика электронной почты. -->
    Name="MailCaptureLine"
    Packer="MailCapturePacker"
    Sender="AnyProgramSender2"
    WorkingFolder="C:\DCS\MailCapture\MailCaptureLineTemp"/>
</Lines>
```

Папки для временных файлов DirectumRX

Чтобы настроить папку для хранения временных файлов, создаваемых при отправке в DirectumRX захваченных документов, в файле CaptureService.exe.config в секции **ProgramSender** укажите значения параметров:

- **WorkingFolder** – путь до папки в которой модуль отправки DCS размещает свои временные файлы;
- **Program** – полный путь до файла SendToRX.bat;
- **ProgramWorkingFolder**– путь до папки, в которой DirectumRX размещает свои временные файлы.

Пример:

```
<ProgramSender
Type="Directum.Dcs.Capture.Senders.ProgramSender.ProgramSenderFactory">
<clear/>
<add
<!-- Настройка модуля отправки для документов из файловой системы. -->
Name="AnyProgramSender"
WorkingFolder="C:\DCS\SomeCapture\AnyProgramSender"
Program="C:\DCS\SendToRX.bat"
ProgramArguments="FileCaptureLine"
ProgramWorkingFolder="C:\DCS\AnyProgramSender\AnyProgramSender_Temp"
...
<add
<!-- Настройка модуля отправки для документов с ящика эл. почты. -->
Name="AnyProgramSender2"
WorkingFolder="C:\DCS\SomeCapture\AnyProgramSender2"
Program="C:\DCS\SendToRX2.bat"
ProgramArguments="MailCaptureLine"
ProgramWorkingFolder="C:\DCS\AnyProgramSender\AnyProgramSender2_Temp"
...
</ProgramSender>
```

Для детальной настройки способа отправки документов в DirectumRX измените значения остальных параметров секции. Подробнее описание параметров см. в инструкции по установке DCS, раздел «Настройка модуля отправки». Входит в комплект поставки DCS.

Приложение 3. Пример адаптации решения

Постановка задачи

Необходимо реализовать интеллектуальную обработку и занесение в DirectumRX документов, подтверждающих расходы на командировки, например, авиа- и железнодорожные билеты. Документы должны поступать в систему по электронной почте, а сервисы интеллектуальной обработки DIRECTUM Arjo извлекать из них информацию о сумме и валюте документа.

Порядок обработки подтверждающих документов должен быть следующим:

1. Сотрудник фотографирует или сканирует подтверждающие документы и отправляет на выделенный ящик электронной почты.
2. Служба ввода документов DCS захватывает документы, поступившие на выделенный ящик электронной почты.
3. С помощью сервисов DIRECTUM Arjo для каждого документа:
 - распознается текст;
 - определяется вид: авиабилет или железнодорожный билет;
 - из текста извлекаются сумма и валюта документа.
4. В DirectumRX создаются подтверждающие документы, в карточках которых автоматически заполняется вид (авиабилет или железнодорожный билет), сумма и валюта.
Если сервисам DIRECTUM Arjo не удалось определить тип документа, то в DirectumRX создается простой документ.
5. По адресу электронной почты, с которого поступили документы, определяется сотрудник-отправитель документов.
6. Создается авансовый отчет, который связывается с комплектом подтверждающих документов.
7. Поступивший комплект подтверждающих документов вместе с авансовым отчетом отправляется задачей на проверку сотруднику-отправителю документов. При этом в карточках документов наглядно выделены цветом распознанные факты и поля, заполненные автоматически.
8. В рамках задачи сотрудник-отправитель документов проверяет полноту комплекта, заполняет необходимые данные в карточках подтверждающих документов, заполняет авансовый отчет. После этого весь комплект документов вместе с авансовым отчетом отправляет на согласование по регламенту.

ВАЖНО. В данном примере рассмотрим создание подтверждающих документов, реализацию классификации, распознавания фактов и заполнения полей в карточках. Для упрощения в данном примере определение сотрудника по адресу электронной почты и создание авансового отчета не рассматривается.

Этап 1. Создание типов документов

На данном этапе в среде разработки DirectumRX необходимо создать новый тип документов **Подтверждающий документ** с видами **Авиабилет** и **Железнодорожный билет**. В карточке нового типа документов должны быть поля для указания суммы и валюты поступившего документа.

Для этого:

1. В среде разработки DirectumRX создайте решение CustomSmartCapture. Подробнее см. в справке по среде разработки DirectumRX, раздел [«Решение»](#).
2. В новом решении добавьте создайте модуль CustomCapture. Подробнее см. в справке по среде разработки DirectumRX, раздел [«Модуль»](#).
3. В новом модуле создайте тип документа **Подтверждающий документ** (SupportingDocuments). В качестве базового типа сущности укажите тип IOfficialDocument. Подробнее см. в справке по среде разработки DirectumRX, раздел [«Типы сущностей»](#).
4. В редакторе форм в карточку нового типа документа добавьте поля **Сумма** (Amount) и **Валюта** (Currency). Подробнее см. в справке по среде разработки DirectumRX, разделы [«Формы»](#) и [«Контроль»](#).
5. В инициализации модуля CustomCapture для нового типа документа добавьте логику создания видов по умолчанию **Авиабилет** и **Железнодорожный билет**.

Перед этим создайте тип документа **Подтверждающий документ** в служебном справочнике **Типы документов**. Записи данного справочника создаются только программно.

Для создания типов и видов документов используйте функции, определенные в серверном коде модуля Docflow – **CreateDocumentType()** и **Create DocumentKind()** соответственно.

Подробнее см. в справке по среде разработки DirectumRX, раздел [«Логика инициализации модуля»](#).

Этап 2. Классификация документов и извлечение фактов

Для извлечения фактов из авиа- и железнодорожных билетов в сервисах DIRECTUM Ario есть преднастроенные правила airticket и railwayticket.

На данном этапе разработки:

- задается соответствие вида документа в DirectumRX и правила извлечения фактов DIRECTUM Ario;
- вида документа в DirectumRX и кода, обрабатывающего результат распознавания;

Для этого:

1. В среде разработки DirectumRX переключите модуль Capture решения [Sungero.SmartCapture](#). Подробнее см. в справке по среде разработки DirectumRX, раздел [«Пример переключения модуля»](#).
2. Чтобы задать соответствие типа документа DirectumRX и правила извлечения фактов DIRECTUM Ario, в переключенном модуле Capture переключите клиентскую функцию **GetClassRuleMapping**:

```
public override System.Collections.Generic.Dictionary<string, string>
GetClassRuleMapping()
{
    var mapping = base.GetClassRuleMapping();
    mapping.Add("Авиабилет", "airticket");
    mapping.Add("Железнодорожный билет", "railwayticket");
    return mapping;
}
```


3. Чтобы задать соответствие типа документа в DirectumRX и кода обработки результата распознавания, в перекрытом модуле Capture перекройте серверную функцию **CreateDocumentByRecognizedDocument**:

```
public virtual IOfficialDocument CreateRailwayTicket(Sungero.Capture.Structures.  
Module.IRecognitionResult recognizedDocument, string fromEmail)  
{  
    var document = CustomCapture.SupportingDocuments.Create();  
    var props = document.Info.Properties;  
  
    // Заполнить вида документа.  
    document.DocumentKind =  
    Docflow.PublicFunctions.DocumentKind.GetNativeDocumentKind(Sungero.CustomCapture  
    .PublicConstants.Module.Initialize.RailwayTicketKindGuid);  
  
    var facts = recognizedDocument.Facts;  
    var amountFacts = GetOrderedFacts(facts, "DocumentAmount", "Amount");  
    var amountFact = amountFacts.FirstOrDefault();  
    if (amountFact != null)  
    {  
        // Заполнить сумму.  
        document.Amount = GetFieldNumericalValue(amountFact, "Amount");  
        LinkFactAndProperty(recognizedDocument, amountFact, "Amount", props.Amount.N  
ame, document.Amount);  
  
        // Заполнить валюту.  
        var currencyCode = GetFieldValue(amountFact, "Currency");  
        document.Currency = Commons.Currencies.GetAll(x => x.NumericCode ==  
currencyCode).FirstOrDefault();  
        LinkFactAndProperty(recognizedDocument, amountFact, "Currency", props.Curren  
cy.Name, document.Currency);  
    }  
  
    return document;  
}  
  
public virtual IOfficialDocument CreateAirTicket(Sungero.Capture.  
Structures.Module.IRecognizedDocument recognizedDocument, string fromEmail)  
{  
    var document = CustomCapture.SupportingDocuments.Create();  
  
    document.DocumentKind =  
    Docflow.PublicFunctions.DocumentKind.GetNativeDocumentKind(Sungero.CustomCapture  
    .PublicConstants.Module.Initialize.AirTicketKindGuid);  
    // Остальной программный код аналогичен функции CreateRailwayTicket.  
    return document;  
}
```

В результате на основе полученной информации о типе документа и фактах, извлеченных сервисами DIRECTUM Ario, в DirectumRX создается новый подтверждающий документ, с заполненным видом (авиабилет или железнодорожный билет), суммой и валютой.

ПРИМЕЧАНИЕ. Для упрощения в данном примере остальные поля карточки не заполняются.

Чтобы проверить результат разработки, обучите сервисы DIRECTUM Ario на реальных авиа- и железнодорожных билетах:

1. [Подготовьте документы для обучения](#). Потребуется минимум 100 документов обоих типов. На сервере, где установлены сервисы DIRECTUM Ario, в сетевой папке создайте структуру папок документов каждого типа.
2. [Обучите классификатор](#) на подготовленных документах.
3. [Проверьте захват документов](#).

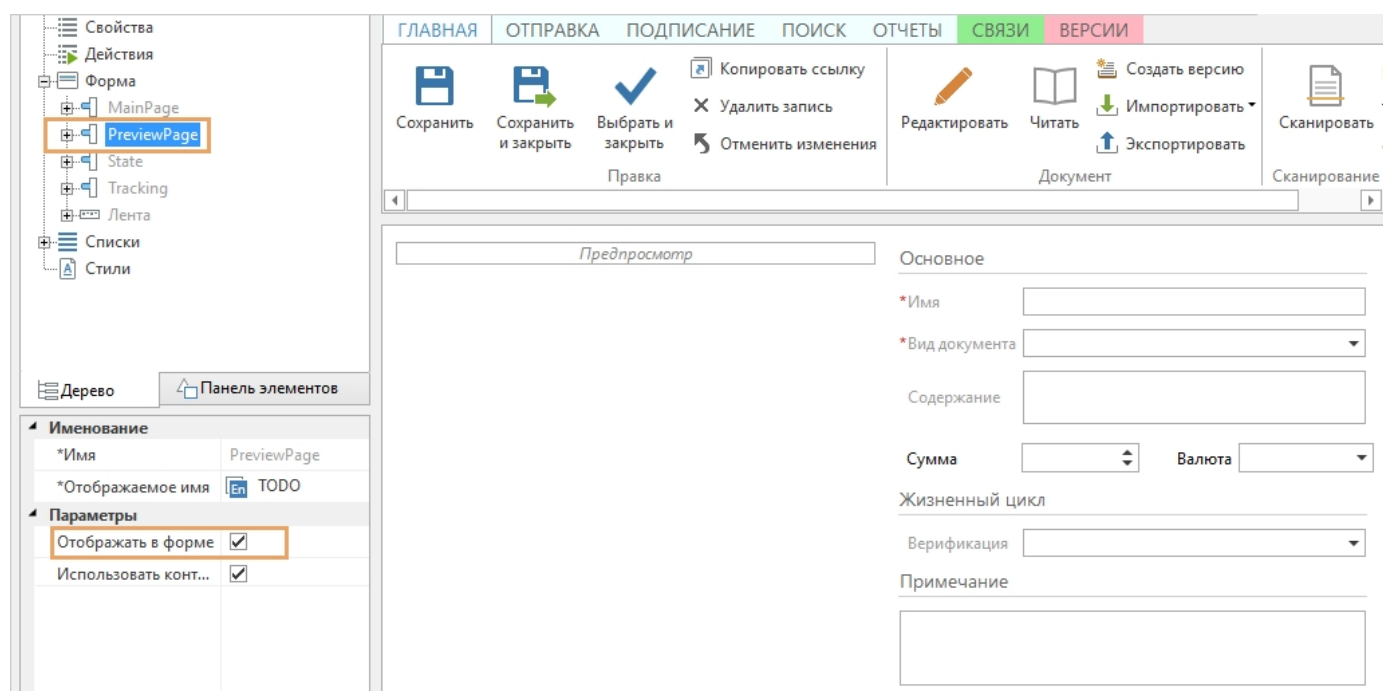
Этап 3. Режим верификации

На данном этапе реализуется наглядное выделение распознанных фактов и заполненных полей в карточках документов. Необходимо, чтобы:

- пока документ находится на стадии верификации (в поле **Верификация** указано значение **В процессе**), при открытии карточки документа отображалась вкладка с областью предпросмотра содержимого, а распознанные факты выделялись цветом в тексте документа и полях карточки;
- после завершения верификации (в поле **Верификация** указано значение **Завершена**), вкладка с областью предпросмотра скрывалась, а распознанные факты в полях карточки не подсвечивались.

Для этого:

1. Для нового типа документов SupportingDocuments в редакторе типа сущности перейдите в узел «Форма», в дереве элементов выберите вкладку **PreviewPage** и на панели свойств установите флажок **Отображать на форме**:



2. Настройте отображение поля **Верификация**. Для этого:
 - в редакторе типа сущности в узле «Форма», последовательно выберите **PreviewPage**, **Verification** и на панели свойств установите флажок **Отображать на форме**;
 - в редакторе типа сущности перейдите в узел «Свойства», в дереве элементов выберите свойство **Verification** и установите флажок **Отображать в карточке**.

3. В событиях вкладки «Показ формы» и «Обновление формы» добавьте код для активации и скрытия вкладки **PreviewPage** в зависимости от значения поля **Верификация**:

```
public override void Showing(Sungero.Presentation.FormShowingEventArgs e)
{
    base.Showing(e);
    Sungero.Capture.PublicFunctions.Module.SwitchVerificationMode(_obj);
}

public override void Refresh(Sungero.Presentation.FormRefreshEventArgs e)
{
    base.Refresh(e);
    Sungero.Capture.PublicFunctions.Module.SwitchVerificationMode(_obj);
}
```