

DirectumRX 3.3. Инструкция по установке сервисов Ario

Содержание

Общие сведения	2
Типовые требования к серверам	2
TextExtractor Service.....	3
SmartService, ClassifierService, FactExtractor Service и Object Detection Service	4
Установка сервисов	5
Службы Internet Information Services	12
Установка служб IIS 8.5	13
Установка служб IIS 10	15
Настройка сервисов	17
Конфигурационные файлы сервисов Ario	17
SmartService	17
TextExtractor Service	19
Classifier Service	20
FactExtractor Service	22
Object Detection Service	23
Справочная информация	23
Список извлекаемых фактов	24
Сторонние компоненты	28

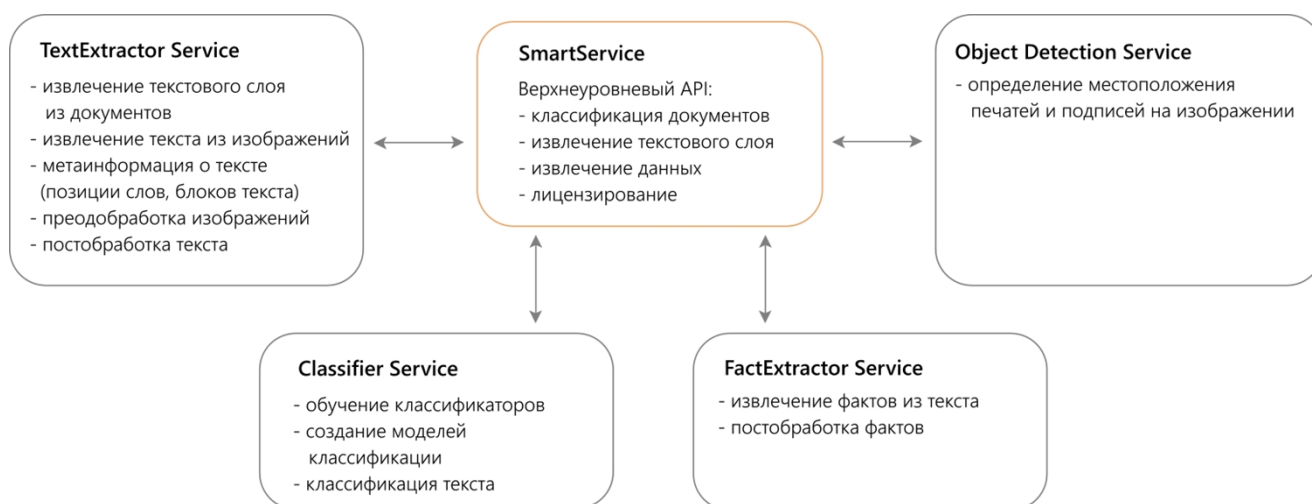
Общие сведения

Система сервисов Arío, основанная на инструментах искусственного интеллекта, автоматизирует рутинные операции в DirectumRX. Сервисы Arío обеспечивают обработку потока документов для их классификации, извлечения значимых данных, автоматической маршрутизации и регистрации.

Сервисы обрабатывают как объекты, содержащие текст, так и документы без текстового слоя:

- извлекают текстовый слой из изображений (jpg, jpeg, png, bmp, gif, tif, tiff, pdf) и текстовых документов (doc, docx, dot, dotx, rtf, odt, ott, txt, xls, xlsx, ods, pdf);
- разделяют комплекты на отдельные документы для независимой обработки. При этом не требуют дополнительных средств – белых листов или штрихкодов;
- конвертируют документы в формат pdf с текстовым слоем для поиска по тексту;
- обучают классификаторы на подготовленном наборе данных: документах, записях справочников и т.д.;
- классифицируют документы в разных разрезах. Например, по типам: договоры, служебные записки, акты, приказы и т.д.;
- извлекают из текста нужные данные, например, дату и номер договора, в том числе данные таблиц;
- рассчитывают вероятность правильного извлечения данных;
- определяют наличие и координаты печатей и подписей на документах.

Сервисы Arío:



Описание методов API сервисов Arío, примеры запросов и ответов доступны по [ссылке](#).

Типовые требования к серверам

Требования к серверам зависят от объема документов, который обрабатывается сервисами Arío в пиковые нагрузки.

Рекомендуется устанавливать сервис [TextExtractor Service](#) на отдельный компьютер. Сервисы [SmartService](#), [Classifier Service](#), [FactExtractor Service](#) и [Object Detection Service](#), как правило, устанавливаются на один компьютер.

Тестирование сервисов Arjo проводилось на указанных версиях программных компонентов. При использовании более новых версий работоспособность сервисов не гарантируется.

Прочие требования:

- [СУБД PostgreSQL 9.6.3](#);
- служба ввода документов, если планируется обрабатывать документы со сканера и электронной почты;
- служба интеграции с системами обмена, если планируется обрабатывать документы из систем обмена.

TextExtractor Service

Компьютер, на котором устанавливается сервис TextExtractor Service, должен удовлетворять требованиям:

Компонент	Требование
Процессор (Intel/AMDсовместимый x64)	Высокочастотный с поддержкой Hyper-Threading
Память (ОЗУ)	Количество аппаратных потоков * 3 ГБ (см. расчет ниже)
Сетевой адаптер	1 Гбит/с и выше
ОС	Microsoft Windows Server 2012 R2/2016 При использовании ОС Microsoft Windows Server 2012 R2 дополнительно необходимо включить и настроить роль сервера Режим Windows Server Essentials , затем установить пакет мультимедиа для Windows Server Essentials
Прочее	<ul style="list-style-type: none">• Microsoft Internet Information Services 7.5 и выше• Microsoft .NET Framework 4.7.2 и выше• Распространяемый компонент Microsoft Visual C++ для Visual Studio 2017 (vc_redist.x64.exe)

Основное масштабирование обеспечивается за счет увеличения мощности или количества компьютеров, где установлен сервис.

Максимальный объем страниц, обрабатываемых в единицу времени, линейно зависит от количества аппаратных потоков станций извлечения текстового слоя и от производительности системы хранения временных данных. Определите необходимую конфигурацию оборудования по таблице (в расчетах день учитывается как 24 часа):

Количество страниц в час (день)	Количество аппаратных потоков	Объем оперативной памяти, ГБ
< 500 (12000)	4	12
500 (12000) – 1000 (24000)	8	24
1000 (24000) – 4000 (96000)	32	96
4000 (96000) – 10000 (240000)	80	240
10000 (240000) – 16000 (384000)	128	384

Объем оперативной памяти рассчитывается по формуле: количество потоков * 3 ГБ.

Процесс извлечения текстового слоя в большей части зависит от скорости CPU и производительности системы хранения временных данных. Архитектура системы подразумевает, что временные документы должны храниться на ресурсе, одновременно доступном для всех станций распознавания текста (в случае горизонтального масштабирования). На этом ресурсе рекомендуется использовать высокоскоростные HDD (SSD), объединенные в конфигурации RAID, позволяющие увеличить скорость чтения и записи.

Предпочтительно использовать в качестве станций извлечения текстового слоя один мультипроцессорный сервер с высокопроизводительной системой хранения данных. В этом случае минимизируются затраты на передачу временных файлов по сети между станциями обработки и ресурсом хранения данных.

Объем системы хранения данных можно рассчитать по формуле:

*Объем системы хранения = количество обрабатываемых страниц в день * средний размер страницы (МБ) * период хранения временных документов.*

Например, для 10000 обрабатываемых страниц в день со средним размером страницы 2 МБ и при периоде хранения временных документов (по умолчанию) 72 часа (3 дня), объем системы хранения составит: $10000 * 2 \text{ МБ} * 3 = 60 \text{ ГБ}$.

SmartService, ClassifierService, FactExtractor Service и Object Detection Service

Компьютер, на который устанавливаются сервисы SmartService, ClassifierService, FactExtractor Service и Object Detection Service, должен соответствовать следующим минимальным требованиям:

Компонент	Требование
Процессор (Intel/AMD-совместимый x64)	64-х разрядный процессор с частотой 2,4 ГГц или выше (4 аппаратных потока или более)
Память (ОЗУ)	8 ГБ
Свободное место (HDD)	Установка: 3 ГБ Временные файлы в процессе работы: от 5 ГБ
Сетевой адаптер	1 Гбит/с
ОС	Microsoft Windows Server 2012 R2/2016 При использовании ОС Microsoft Windows Server 2012 R2 дополнительно необходимо включить и настроить роль сервера Режим Windows Server Essentials , затем установить пакет мультимедиа для Windows Server Essentials
Прочее	<ul style="list-style-type: none"> Microsoft Internet Information Services 7.5 и выше (для SmartService) ASP.NET Core/.NET Core: Runtime & Hosting Bundle 2.0.5 и выше

Масштабирование этих сервисов в меньшей степени зависит от объема обрабатываемых документов в единицу времени. Определите необходимую конфигурацию оборудования по таблице (в расчетах день учитывается как 24 часа):

Количество страниц в час (день)	Количество аппаратных потоков	Объем оперативной памяти, ГБ
< 500 (12000)	4	8
500 (12000) – 4000 (96000)	8	16
4000 (96000) – 10000 (240000)	16	32
10000 (240000) – 16000 (384000)	24	48

Для более гибкого масштабирования сервисы можно установить на отдельных компьютерах.

Установка сервисов

Перед установкой сервисов Ario:

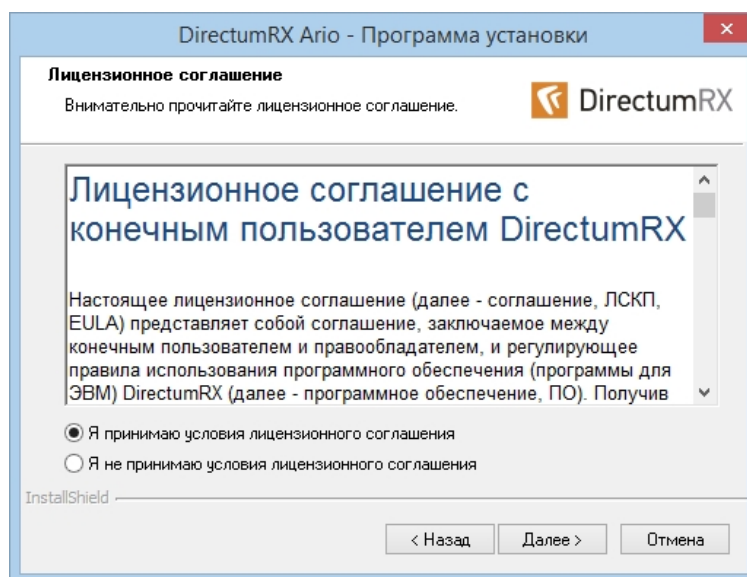
1. Ознакомьтесь с [типовыми требованиями](#) к аппаратному и программному обеспечению и скачайте указанные версии программных компонентов, необходимых для работы сервисов;
2. В указанном порядке установите компоненты:
 - Microsoft .NET Framework;
 - распространяемый компонент Microsoft Visual C++ для Visual Studio 2017 (vc_redist.x64.exe);
 - СУБД PostgreSQL;
 - [Microsoft Internet Information Services](#). После установки перезагрузите компьютер;
 - ASP.NET Core/.NET Core: Runtime & Hosting Bundle;
3. Запросите ключ активации сервисов в службе поддержки DirectumRX и положите полученный файл ActivationKey.aak в папку с программой установки сервисов Ario. После установки и запуска сервисов активация выполнится автоматически.

При необходимости сервисы можно активировать и после завершения установки. Для этого положите файл в папку с установленным сервисом Smart Service, например, C:\inetpub\Ario\SmartService и запустите страницу для проверки состояния сервисов. Активация выполнится автоматически.

ПРИМЕЧАНИЕ. Файл активации имеет ограниченный период действия. Активацию необходимо выполнить в период, указанный службой поддержки. Для переактивации сервисов, например, при истечении срока действия лицензии или изменении конфигурации оборудования, запросите ключ активации в службе поддержки DirectumRX, удалите файл LicenseKey.alk из папки с установленным сервисом Smart Service и вновь выполните активацию.

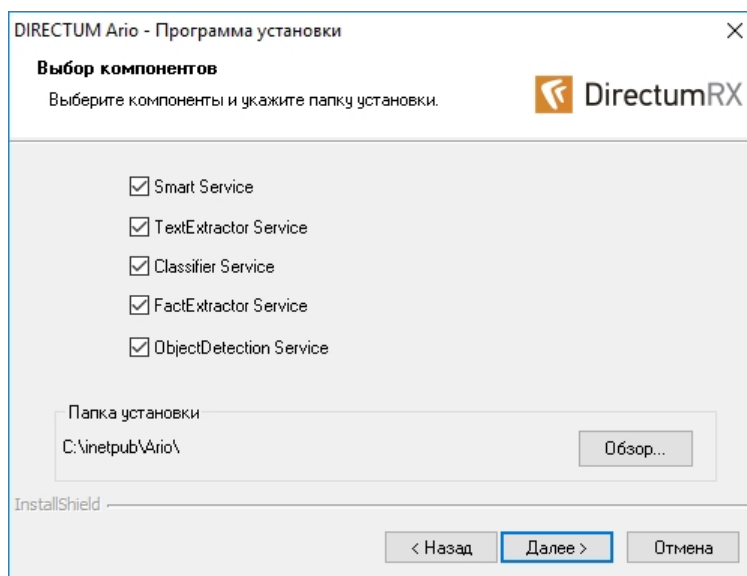
Запустите программу установки сервисов Ario.

1. В окне «Лицензионное соглашение»:



- ознакомьтесь с текстом лицензионного соглашения;
- установите переключатель **Я принимаю условия лицензионного соглашения**;
- нажмите на кнопку **Далее >**.

2. В окне «Выбор компонентов»:



- установите флажки напротив компонентов, которые необходимо установить. Сервис Text Extractor Service рекомендуется устанавливать на отдельный компьютер;
ПРИМЕЧАНИЕ. В зависимости от выбранных компонентов некоторые этапы установки могут пропускаться.
- укажите папку для размещения файлов сервисов. Значение по умолчанию C:\inetpub\Ario;
- нажмите на кнопку **Далее >**.

3. В окне «Параметры веб-сайта Smart Service»:

DIRECTUM Ario - Программа установки

Параметры веб-сайта Smart Service

Укажите параметры создаваемого веб-сайта.

*Адрес веб-сайта: SMART

*Порт: 61100

Внимание! Если имя компьютера не используется в качестве адреса веб-сайта, укажите веб-сайт в DNS или hosts-файлах компьютеров сети.

InstallShield

< Назад **Далее >** Отмена

- в поле ***Адрес веб-сайта** укажите имя создаваемого сайта сервиса Smart Service. По умолчанию заполняется именем текущего компьютера;
- в поле ***Порт** укажите порт создаваемого веб-сайта. Значение по умолчанию **61100**;
- нажмите на кнопку **Далее >**.

4. В окне «Параметры веб-сайта Text Extractor»:

DIRECTUM Ario - Программа установки

Параметры веб-сайта Text Extractor

Укажите параметры создаваемого веб-сайта.

*Адрес веб-сайта: EXTRACTOR

*Порт: 9027

Внимание! Если имя компьютера не используется в качестве адреса веб-сайта, укажите веб-сайт в DNS или hosts-файлах компьютеров сети.

InstallShield

< Назад **Далее >** Отмена

- в поле ***Адрес веб-сайта** укажите имя создаваемого сайта сервиса Text Extractor Service. По умолчанию заполняется именем текущего компьютера;
- в поле ***Порт** укажите порт создаваемого веб-сайта. Значение по умолчанию **9027**;
- нажмите на кнопку **Далее >**.

5. В окне «Пользователь пула приложений Smart Service»:

- выберите учетную запись, от имени которой будет работать пул приложений сервиса Smart Service.

Использовать системную учетную запись. Возможные значения: **Application Pool Identity, Network Service, Local Service, Local System.**

Использовать указанную учетную запись. В этом случае заполните поля ***Имя пользователя** и **Пароль**.

Рекомендуется использовать учетную запись **Application Pool Identity**;

- нажмите на кнопку **Далее >**.

6. В окне «Пользователь пула приложений Text Extractor»:

- выберите учетную запись, от имени которой будет работать пул приложений сервиса Text Extractor Service.

Использовать системную учетную запись. Возможные значения: **Application Pool Identity, Network Service, Local Service, Local System.**

Использовать указанную учетную запись. В этом случае заполните поля ***Имя пользователя** и **Пароль**.

Рекомендуется использовать учетную запись **Application Pool Identity**;

- нажмите на кнопку **Далее >**.

7. В окне «Подключение к базе данных Smart Service»:

DIRECTUM Ario - Программа установки

Подключение к базе данных SmartService

Укажите параметры подключения.

*Адрес: localhost

*Порт: 5432

*Логин: postgres

Пароль:

*Имя базы данных: Ario_Smart_Service

InstallShield

< Назад **Далее >** Отмена

- в поле ***Адрес** укажите адрес СУБД PostgreSQL. Значение по умолчанию **localhost**;
- в поле ***Порт** укажите порт для подключения к СУБД. Если СУБД установлена локально, то значение по умолчанию берется из ветки реестра HKEY_LOCAL_MACHINE\SOFTWARE\PostgreSQL\Services\postgresql-x64-9.6\Port. Иначе подставляется значение **5432**;
- в полях ***Логин** и **Пароль** укажите логин и пароль для подключения к СУБД;
- в поле ***Имя базы данных** укажите имя ранее созданной базы данных Smart Service или задайте имя новой базы данных. В последнем случае база данных Smart Service будет создана при первом запуске сервисов;
- нажмите на кнопку **Далее >**.

8. В окне «Параметры Classifier Service»:

DIRECTUM Ario - Программа установки

Параметры Classifier Service

Заполните параметры сервиса.

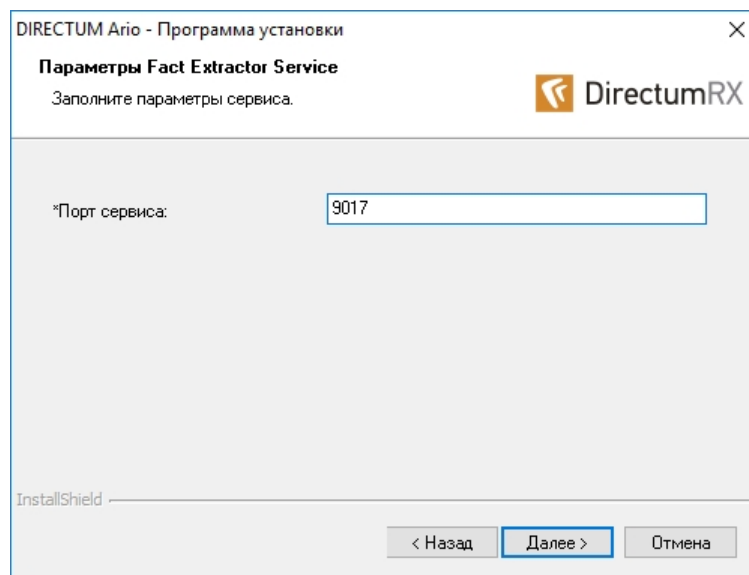
*Порт сервиса: 9007

InstallShield

< Назад **Далее >** Отмена

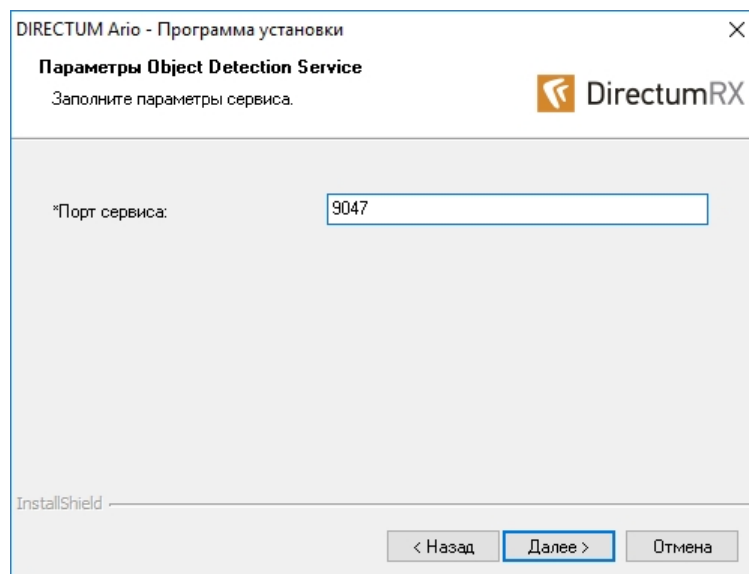
- в поле ***Порт сервиса** укажите порт устанавливаемого сервиса классификации. Значение по умолчанию **9007**;
- нажмите на кнопку **Далее >**.

9. В окне «Параметры Fact Extractor Service»:



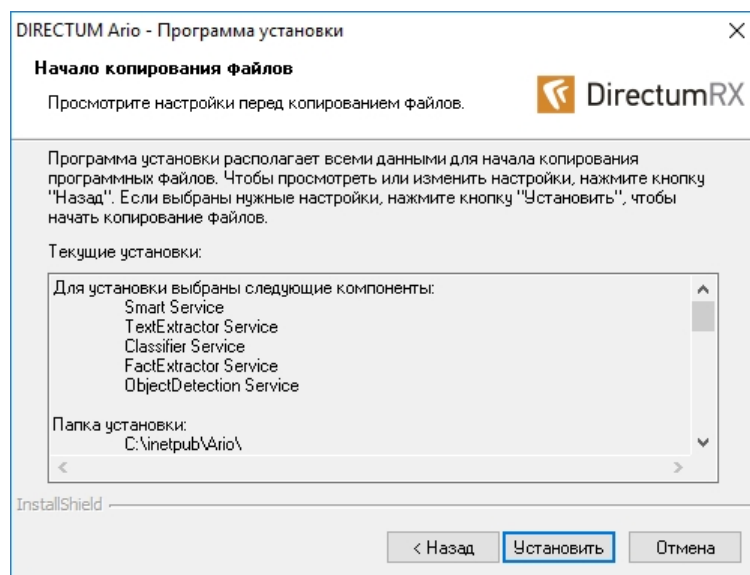
- в поле ***Порт сервиса** укажите порт устанавливаемого сервиса извлечения фактов. Значение по умолчанию **9017**;
- нажмите на кнопку **Далее >**.

10. В окне «Параметры Object Detection Service»:



- в поле ***Порт сервиса** укажите порт устанавливаемого сервиса поиска объектов на изображениях. Значение по умолчанию **9047**;
- нажмите на кнопку **Далее >**.

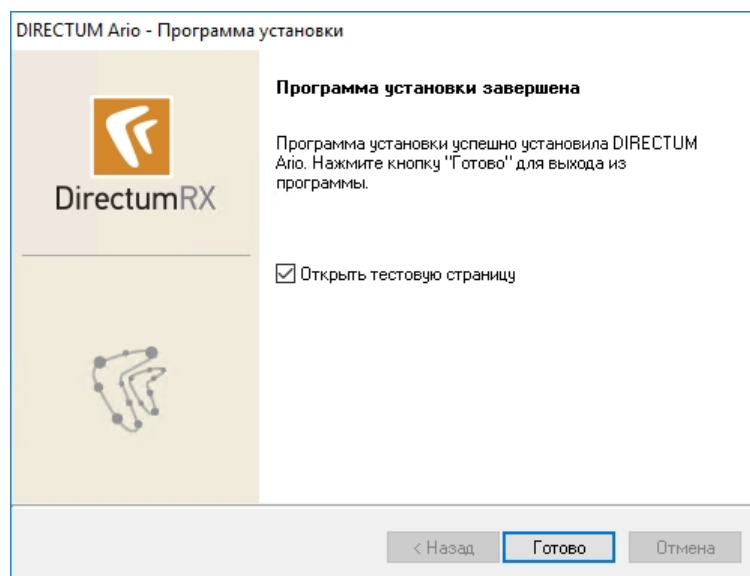
11. В окне «Начало копирования файлов»:



- ознакомьтесь с параметрами установки;
- нажмите на кнопку **Установить**.

В результате установки отобразится информация об успешном завершении или возникших ошибках.

12. в окне «Программа установки завершена»:



- оставьте флажок **Открыть тестовую страницу**, чтобы в браузере открылась страница с текущим состоянием сервисов. Состояние **is running** означает, что сервис работает;
- нажмите **Готово**.

В результате установки сервисов Ario:

1. В файловой системе:

- добавляется папка с файлами сервисов Ario. По умолчанию C:\inetpub\Ario;
- добавляется папка %ProgramData%\Python, которая содержит файлы Python.

2. В структуре IIS:
 - создаются пулы приложений «SmartServiceAppPool» и «TextExtractorServiceAppPool»;
 - добавляются новые веб-сайты «SmartService» и «TextExtractorService»;
 - пользователям для запуска пулов приложений назначаются необходимые [права](#) на папки.
3. В список служб Windows добавляются службы Classifier Service, FactExtractor Service и Object Detection Service.

Службы Internet Information Services

Службы Internet Information Services (далее IIS) являются компонентами Windows, облегчающими публикацию информации и внедрение деловых приложений в Интернете. IIS упрощает создание платформы для сетевых приложений.

В среднем установка службы Internet Information Services занимает 10-15 минут.

Перед началом установки IIS необходимо войти в Windows Server 2012R2/2016 от имени пользователя, обладающего правами администратора. Рекомендуется использовать пользователя, в имени которого нет русских букв.

Порядок установки служб:

- [IIS 8.5](#) (Microsoft Windows Server 2012 R2);
- [IIS 10](#) (Microsoft Windows Server 2016).

Компоненты IIS можно установить через командную строку:

1. Смонтировать образ диска текущей операционной системы в виртуальный привод (требуется, так как по умолчанию ОС Microsoft Windows Server 2012 и выше не включает Microsoft.NET Framework 3.5).
2. Запустите CMD.exe и выполните команду:

```
DISM.EXE /enable-feature /Source:<Буква диска, на который был смонтирован образ ОС, например, E:>\Sources\sxs /online /featureName:IIS-WebServerRole /featureName:IIS-WebServer /featureName:IIS-CommonHttpFeatures /featureName:IIS-StaticContent /featureName:IIS-DefaultDocument /featureName:IIS-DirectoryBrowsing /featureName:IIS-HttpErrors /featureName:IIS-CGI /featureName:IIS-ISAPIExtensions /featureName:IIS-ISAPIFilter /featureName:IIS-Security /featureName:IIS-BasicAuthentication /featureName:IIS-WindowsAuthentication /featureName:IIS-RequestFiltering /featureName:IIS-Performance /featureName:IIS-HttpCompressionStatic /featureName:IIS-HealthAndDiagnostics /featureName:IIS-HttpLogging /featureName:IIS-RequestMonitor /featureName:IIS-HttpTracing /featureName:IIS-ASPNET45 /featureName:IIS-ManagementScriptingTools /featureName:IIS-ManagementService /featureName:WCF-HTTP-Activation45 /featureName:Windows-Identity-Foundation /featureName:IIS-ASPNET /featureName:IIS-ASP /featureName:IIS-ISAPIExtensions /featureName:IIS-ISAPIFilter /featureName:IIS-NetFxExtensibility /featureName:IIS-NetFxExtensibility45 /featureName:IIS-NetFxExtensibility /featureName:NetFx3 /featureName:NetFx3ServerFeatures /featureName:NetFx4Extended-ASPNET45 /featureName:IIS-ApplicationDevelopment /featureName:NetFx3 /all /LimitAccess
```

Установка служб IIS 8.5

1. Запустите приложение **Диспетчер серверов**. Например, на начальной странице щелкните заголовок **Диспетчер сервера**.
2. На панели мониторинга диспетчера серверов в меню **Управление** выберите пункт **Добавить роли и компоненты**. Запустится мастер добавления ролей и компонентов.
3. На странице «Перед началом работы» нажмите на кнопку **Далее**.
4. На странице «Выбор типа установки» выберите пункт **Установка ролей или компонентов** и нажмите на кнопку **Далее**.
5. На странице «Выбор целевого сервера»:
 - выберите пункт **Выбор сервера из пула серверов**;
 - укажите сервер, на котором необходимо добавить роли и компоненты;
 - нажмите на кнопку **Далее**.
6. На странице «Выбор ролей сервера» установите флажок **Веб-сервер (IIS)**.
7. В открывшемся окне «Добавить компоненты, которые требуются для Веб-сервер (IIS)» нажмите на кнопку **Добавить компоненты**, если требуется установить консоль управления IIS.
8. На странице «Выбор компонентов»:
 - установите флажки **Функции .NET Framework 4.5/Службы WCF/Активация по HTTP** и **Windows Identity Foundation 3.5**;
 - в открывшемся окне «Добавить компоненты, необходимые для Активация по HTTP?» нажмите на кнопку **Добавить компоненты**. В результате будут автоматически установлены флажки **.NET Framework 4.5** и **ASP.NET 4.5** вместе с другими связанными флажками;
 - нажмите на кнопку **Далее**.
9. На странице «Роль веб-сервера (IIS)» нажмите на кнопку **Далее**.

10. На странице «Выбор служб ролей»:

- укажите службы роли:

☒ Веб-сервер

▾ ☒ Безопасность

☒ Фильтрация запросов

☐ IP-адрес и ограничения домена

☐ Авторизация URL-адреса

☐ Дайджест-проверка подлинности

☒ Обычная проверка подлинности

☒ Проверка подлинности Windows

☐ Проверка подлинности с сопоставлением сертификата клиента

☐ Проверка подлинности с сопоставлением сертификата клиента IIS

☐ Централизованная поддержка SSL-сертификата

▾ ☒ Исправность и диагностика

☒ Ведение журнала HTTP

☐ Ведение журнала ODBC

☒ Монитор запросов

☐ Настраиваемое ведение журнала

☐ Средства ведения журнала

☒ Трассировка

▾ ☒ Общие функции HTTP

☒ Документ по умолчанию

☒ Обзор каталога

☒ Ошибки HTTP

☒ Статическое содержимое

☒ Перенаправление HTTP

☐ Публикация WebDAV

▾ ☒ Производительность

☒ Сжатие статического содержимого

☐ Сжатие динамического содержимого

▾ ☒ Разработка приложений

☒ ASP

☒ ASP.NET 3.5

☒ ASP.NET 4.5

☐ CGI

☐ Включения на стороне сервера (SSI)

☐ Инициализация приложений

☐ Протокол WebSocket

☒ Расширения ISAPI

☒ Расширяемость .NET 3.5

☒ Расширяемость .NET 4.5

☒ Фильтры ISAPI

☐ FTP-сервер

☐ Служба FTP

☐ Расширяемость FTP

☐ Внутрипроцессное веб-ядро IIS

☒ Средства управления

☒ Консоль управления службами IIS

☒ Наборы символов и средства управления службами IIS

☒ Служба управления

- нажмите на кнопку **Далее**.

11. На странице «Подтверждение выбранных элементов для установки» нажмите на кнопку **Установить**.12. На странице «Ход выполнения установки» убедитесь, что установка роли веб-сервера (IIS) и требуемых служб ролей успешно завершена, нажмите на кнопку **Заккрыть**.

13. При установке IIS автоматически создается предопределенный пользователь IUSR. Для него необходимо настроить права на изменение каталога временных файлов Windows: %SYSTEMROOT%\Temp.

Установка служб IIS 10

1. В меню **Пуск** выберите пункт **Диспетчер серверов**.
2. На панели мониторинга диспетчера серверов в меню **Управление** выберите пункт **Добавить роли и компоненты**. Запустится мастер добавления ролей и компонентов.
3. На странице «Перед началом работы» нажмите на кнопку **Далее**.
4. На странице «Выбор типа установки» выберите пункт **Установка ролей или компонентов** и нажмите на кнопку **Далее**.
5. На странице «Выбор целевого сервера»:
 - выберите пункт **Выберите сервер из пула серверов**;
 - укажите сервер, на котором необходимо добавить роли и компоненты;
 - нажмите на кнопку **Далее**.
6. На странице «Выбор ролей сервера» установите флажок **Веб-сервер (IIS)**.
7. В открывшемся окне «Добавить компоненты, необходимые для Веб-сервер (IIS)?» нажмите на кнопку **Добавить компоненты**.
8. На странице «Выбор ролей сервера» нажмите на кнопку **Далее**.
9. На странице «Выбор компонентов»:
 - установите флажки **Функции .NET Framework 4.6/Службы WCF/Активация по HTTP** и **Windows Identity Foundation 3.5**;
 - в открывшемся окне «Добавить компоненты, необходимые для Активация по HTTP?» нажмите на кнопку **Добавить компоненты**. В результате автоматически установятся флажки **.NET Framework 4.6** и **ASP.NET 4.6** вместе с другими связанными флажками;
 - нажмите на кнопку **Далее**.
10. На странице «Роль веб-сервера (IIS)» нажмите на кнопку **Далее**.

11. На странице «Выбор служб ролей»:

- укажите службы роли:

☒ Веб-сервер

- ☒ Безопасность
 - ☒ Фильтрация запросов
 - ☐ IP-адрес и ограничения домена
 - ☐ Авторизация URL-адреса
 - ☐ Дайджест-проверка подлинности
 - ☒ Обычная проверка подлинности
 - ☒ Проверка подлинности Windows
 - ☐ Проверка подлинности с сопоставлением сертификата клиента
 - ☐ Проверка подлинности с сопоставлением сертификата клиента IIS
 - ☐ Централизованная поддержка SSL-сертификата
- ☒ Исправность и диагностика
 - ☒ Ведение журнала HTTP
 - ☐ Ведение журнала ODBC
 - ☒ Монитор запросов
 - ☐ Настраиваемое ведение журнала
 - ☐ Средства ведения журнала
 - ☒ Трассировка
- ☒ Общие функции HTTP
 - ☒ Документ по умолчанию
 - ☒ Обзор каталога
 - ☒ Ошибки HTTP
 - ☒ Статическое содержимое
 - ☒ Перенаправление HTTP
 - ☐ Публикация WebDAV
- ☒ Производительность
 - ☒ Сжатие статического содержимого
 - ☐ Сжатие динамического содержимого
- ☒ Разработка приложений
 - ☒ ASP
 - ☒ ASP.NET 3.5
 - ☒ ASP.NET 4.7
 - ☐ CGI
 - ☐ Включения на стороне сервера (SSI)
 - ☐ Инициализация приложений
 - ☐ Протокол WebSocket
 - ☒ Расширения ISAPI
 - ☒ Расширяемость .NET 3.5
 - ☒ Расширяемость .NET 4.7
 - ☒ Фильтры ISAPI

☐ FTP-сервер

- ☒ Средства управления
 - ☒ Консоль управления службами IIS
 - ☒ Наборы символов и средства управления службами IIS
 - ☒ Служба управления
- ☐ Управление совместимостью с IIS 6

При установке флажка **ASP.NET 3.5** в открывшемся окне «Добавить компоненты, необходимые для ASP.NET 3.5?» нажмите на кнопку **Добавить компоненты**;

- нажмите на кнопку **Далее**.

12. На странице «Подтверждение установки компонентов» нажмите на ссылку **Указать альтернативный исходный путь**.

13. В открывшемся окне «Указание альтернативного исходного пути»:

- в поле **Путь** укажите расположение исходных файлов для компонентов .NET Framework 3.5. Подробнее см. [статью](#);
- нажмите на кнопку **ОК**.

14. На странице «Подтверждение установки компонентов» нажмите на кнопку **Установить**.15. На странице «Ход установки» убедитесь, что установка роли веб-сервера (IIS) и требуемых служб ролей успешно завершена, а затем нажмите на кнопку **Заккрыть**.

16. При установке IIS автоматически создается предопределенный пользователь IUSR. Для него необходимо настроить права на изменение каталога временных файлов Windows: %SYSTEMROOT%\Temp.

Настройка сервисов

После установки убедитесь, что есть права на изменение папок:

- C:\inetpub\Ario\SmartService – у пользователя, от имени которого работает пул приложений Smart Service;
- C:\inetpub\Ario\TextExtractor – у пользователя, от имени которого работает пул приложений Text Extractor Service.

Конфигурационные файлы сервисов Ario

В процессе работы администратор может редактировать конфигурационные файлы сервисов Ario:

- [SmartService](#)
- [TextExtractor Service](#)
- [Classifier Service](#)
- [FactExtractor Service](#)
- [Object Detection Service](#)

Конфигурационные файлы расположены в папках, в которых [установлены](#) сервисы Ario.

SmartService

Сервис SmartService настраивается в конфигурационном файле appsettings.json.

Структура файла:

```
{
  "ConnectionStrings": {
    "DefaultConnection": "User ID={Имя пользователя};Password={Пароль};
      Host={Адрес};Port={Порт};Database={Имя базы данных};
      Pooling={Использовать пул подключений к БД};"
  },
  "Serilog": {
    "MinimumLevel": {
      "Default": "{Минимальный уровень протоколирования}",
      "Override": {
        "Microsoft.EntityFrameworkCore": "Error",
        "System": "Error"
      }
    },
    "Properties": {
      "Application": "SmartService"
    }
  },
  "SmartServices": {
    "Classification": "{Адрес сервиса классификации}",
    "Training": "{Адрес сервиса обучения классификаторов}",
    "TextExtraction": "{Адрес сервиса извлечения текстового слоя}",
    "FactExtractorService": "{Адрес сервиса извлечения фактов}",
    "AllowLocalPath": {Разрешить обучение классификаторов из локальной папки}
  }
}
```

ConnectionStrings – параметры подключения.

DefaultConnection. Строка подключения к базе данных PostgreSQL, где:

- **UserID** и **Password** – имя и пароль пользователя для подключения к СУБД PostgreSQL;
- **Host** – адрес СУБД PostgreSQL;
- **Port** – порт для подключения к СУБД PostgreSQL;
- **Database** – имя базы данных SmartService;
- **Pooling** – признак использования пула подключений к базе данных. Возможные значения: **True**, **False**. Если установлено значение **True**, то для новых запросов к базе данных используется ранее созданное подключение из пула. Если установлено значение **False**, то каждый раз создается новое подключение. Для ускорения работы сервиса рекомендуется оставить в параметре значение по умолчанию **True**.

Serilog – параметры ведения лог-файлов.

MinimumLevel. Минимальный уровень протоколирования. Значение по умолчанию:

```
"Serilog": {
  "MinimumLevel": {
    "Default": "Information",
    "Override": {
      "Microsoft.EntityFrameworkCore": "Error",
      "System": "Error"
    }
  },
}
```

При указанном значении параметра в лог-файл записываются сообщения о каждом этапе обработки документа, а также ошибки, возникающие при работе сервиса.

Подробнее об изменении значения параметра см. в [документации](#) к библиотеке Serilog.

SmartServices – настройка связи с сервисами Ario.

Classification. Адрес сервиса классификации Classifier Service. Например, **http://localhost:9007**.

Training. Адрес выделенного сервиса обучения классификаторов. Выделенный сервис используется, чтобы обучение не влияло на обработку документов существующими классификаторами. Например, **http://localhost:9007**.

TextExtraction. Адрес сервиса извлечения текстового слоя TextExtractor Service. Например, **http://localhost:9027**.

FactExtractorService. Адрес сервиса извлечения фактов FactExtractor Service. Например, **http://localhost:9017**.

ObjectDetectionService. Адрес сервиса поиска объектов на изображениях Object Detection Service. Например, **http://localhost:9047**.

AllowLocalPath. Разрешить обучение классификаторов из локальной папки сервера. Возможные значения: **True**, **False**. Если установлено значение **False**, повышается безопасность работы. Значение по умолчанию **True**.

TextExtractor Service

Сервис TextExtractor Service настраивается в секции **appSettings** конфигурационного файла web.config. Не рекомендуется изменять остальные секции файла, это может повлиять на работоспособность сервиса.

Структура секции:

```
<appSettings>
  <add key="ExtractionQuality" value="{Качество извлечения текстового слоя}" />
  <!--<add key="TempDocumentsLocation" value="{Путь к папке с временными
документами}"/>-->
  <add key="ClearTempDocumentsOlderThanHours" value="{Интервал удаления
временных документов из папки}" />
  <add key="AlwaysExtractTextFromPdf" value="{Всегда извлекать текстовый слой из
PDF-документов}" />
  <add key="NoTextPdfExtractionDpi" value="{Разрешение для извлечения текста из
PDF-документов}" />
  <add key="UseConvertedPdfFactPositions" value="{Учитывать поворот документов
при получении
координат фактов}" />
  <add key="PageClassifierWordsCount" value="{Количество слов для определения
начала документа}" />
  <add key="UseFixedForms" value="{Использовать жесткие формы документов}" />
  <add key="DebugMode" value="{Включить режим отладки}" />
</appSettings>
```

appSettings – настройки сервиса:

ExtractionQuality. Качество извлечения текстового слоя. Возможные значения:

- **fast.** Высокая скорость извлечения, при этом снижается качество. Значение по умолчанию;
- **quality.** Высокое качество извлечения, при этом снижается скорость.

TempDocumentsLocation. Путь к папке с временными документами, которые создаются в процессе работы сервиса. Параметр используется, если установлено несколько сервисов извлечения текстового слоя, чтобы все они использовали одну и ту же папку. По умолчанию параметр закомментирован.

ClearTempDocumentsOlderThanHours. Интервал удаления временных документов из папки в часах. При перезапуске сервиса из временной папки удаляются документы, созданные ранее этого интервала. Укажите целое число больше нуля. Значение по умолчанию **72**.

AlwaysExtractTextFromPdf. Всегда извлекать текстовый слой из PDF-документов, даже если текстовый слой уже существует. Возможные значения: **True, False**. Значение по умолчанию **True**.

NoTextPdfExtractionDpi. Разрешение, используемое для извлечения текста из PDF-документов без текстового слоя или из документов, у которых игнорируется существующий текстовый слой. Укажите целое число больше нуля. Значение по умолчанию **300**.

UseConvertedPdfFactPositions. Учитывать поворот документов при получении координат фактов. Возможные значения: **True, False**. Укажите значение **False**, если импортируются PDF-документы без предварительной обработки. Если импортируются обработанные PDF-документы, оставьте значение по умолчанию **True**.

PageClassifierWordsCount. Количество слов на странице, которые используются для определения начала нового документа. Укажите целое число больше нуля. Значение по умолчанию **50**.

UseFixedForms. При обработке изображений применять жесткие формы документов, например, формы паспортов РФ. Возможные значения: **True**, **False**. Если нет необходимости в распознавании документов с жесткими формами, для ускорения обработки документов рекомендуется оставить значение по умолчанию **False**.

DebugMode. Включить режим отладки для сохранения промежуточных результатов обработки изображений. Возможные значения: **True**, **False**. Если указано значение **True**, то во временной папке, указанной в параметре **TempDocumentsLocation**, создается папка debug, в которую сохраняются промежуточные результаты обработки изображений на каждом из этапов. Значение по умолчанию **False**.

Classifier Service

Сервис Classifier Service настраивается в конфигурационном файле app.config.

Структура файла:

[DEFAULT]

debug = {Записывать в лог-файл отладочные сообщения}

area = {Используемый раздел конфигурационного файла}

max_models_in_pool = {Максимальное количество моделей классификации в пуле}

max_pool_model_lifetime = {Время хранения модели классификации в пуле}

smart_service_path = {Адрес сервиса SmartService}

dbname = {Имя базы данных PostgreSQL}

dbuser = {Имя пользователя для подключения к СУБД PostgreSQL}

dbhost = {Адрес СУБД PostgreSQL}

dbpass = {Пароль для подключения к СУБД PostgreSQL}

dbport = {Порт для подключения к СУБД PostgreSQL}

csv_text_field = {Название поля для обучения данными в формате CSV}

csv_class_field = {Название поля для обучения данными в формате CSV}

csv_external_id_field = {Название поля для обучения данными в формате CSV}

allow_local_path = {Разрешить обучение из локальной папки}

repository = {Тип репозитория для хранения моделей классификации}

port = {TCP-порт сервиса}

training_service = {Разрешить использовать сервис как сервис обучения классификаторов}

[DEVELOP]

allow_local_path = {Разрешить обучение из локальной папки}

[LOGGING]

period = {Периодичность создания новых лог-файлов}

suffix = {Формат названия лог-файлов}

DEFAULT – основные настройки сервиса:

debug. Записывать в лог-файл отладочные сообщения. Возможные значения: **True**, **False**. Значение по умолчанию **True**.

area. Используемый в данный момент раздел конфигурационного файла. Используется для быстрого переключения между настройками окружения. Если параметр не задан в указанном разделе, используется значение параметра из раздела DEFAULT. Например, **DEVELOP**.

max_models_in_pool. Максимальное количество моделей классификации, хранящихся в невыгружаемом пуле. Если модель классификации находится в пуле, обращение к ней выполняется быстрее, таким образом ускоряется классификация текстов. Укажите целое число больше нуля. Значение по умолчанию **15**.

max_pool_model_lifetime. Время хранения модели классификации в пуле в минутах. Модели, к которым не обращались в течение этого промежутка времени, выгружаются из пула. Укажите целое число больше нуля. Значение по умолчанию **60**.

smart_service_path. Адрес сервиса SmartService. Используется для извлечения текстового слоя из документов при обучении классификатора. Например, **http://localhost:61100**.

dbname. Имя базы данных PostgreSQL.

dbuser. Имя пользователя для подключения к СУБД PostgreSQL.

dbhost. Адрес СУБД PostgreSQL.

dbpass. Пароль для подключения к СУБД PostgreSQL.

dbport. Порт для подключения к СУБД PostgreSQL.

ПРИМЕЧАНИЕ. Параметры подключения к базе данных должны совпадать с параметрами, указанными в [конфигурационном файле](#) SmartService.

csv_text_field, csv_class_field, csv_external_id_field. Названия полей в csv файле для обучения классификаторов через набор данных в формате CSV. Параметры являются внутренними, изменять не рекомендуется.

allow_local_path. Разрешить обучение классификаторов из локальной папки сервера. Возможные значения: **True**, **False**. Если установлено значение **False**, повышается безопасность работы. Значение по умолчанию **True**.

repository. Тип репозитория, используемого для хранения моделей классификации. Параметр является внутренним, изменять не рекомендуется.

port. TCP-порт сервиса. Значение по умолчанию **9007**.

training_service. Разрешить использовать этот сервис в качестве сервиса обучения классификаторов. Возможные значения: **True**, **False**. Значение по умолчанию **True**. Подробнее см. описание параметра **Training** в [конфигурационном файле](#) SmartService.

LOGGING – параметры ведения лог-файлов.

period. Периодичность создания новых лог-файлов. Возможные значения:

- **per_minute** – каждую минуту;
- **per_hour** – каждый час;
- **per_day** – каждый день;
- **midnight** – каждый день в полночь. Значение по умолчанию.

suffix. Формат названия лог-файлов. Значение по умолчанию **%%Y%%m%%d**.

FactExtractor Service

Сервис Classifier Service настраивается в конфигурационном файле app.config.

Структура файла:

[DEFAULT]

debug = {Записывать в лог-файл отладочные сообщения}

area = {Используемый раздел конфигурационного файла}

parser_file_name = {Имя файла синтаксического анализатора для извлечения фактов}

parser_debug = {Записывать в лог-файл отладочные сообщения в процессе извлечения фактов и расчета вероятностей}

parser_pool_length = {Количество экземпляров синтаксического анализатора в пуле}

port = {TCP-порт сервиса}

[LOGGING]

period = {Периодичность создания новых лог-файлов}

suffix = {Формат названия лог-файлов}

DEFAULT – основные настройки сервиса:

debug. Записывать в лог-файл отладочные сообщения. Возможные значения: **True**, **False**. Значение по умолчанию **True**.

area. Используемый в данный момент раздел конфигурационного файла. Используется для быстрого переключения между настройками окружения. Если параметр не задан в указанном разделе, используется значение параметра из раздела **DEFAULT**. Например, **DEVELOP**.

parser_file_name. Имя файла синтаксического анализатора, применяемого для извлечения фактов. Значение по умолчанию **parser.exe**. Параметр является внутренним, изменять не рекомендуется.

parser_debug. Записывать в лог-файл отладочные сообщения, возникающие в процессе извлечения фактов и расчета вероятностей. Возможные значения: **True**, **False**. Значение по умолчанию **False**.

parser_pool_length. Количество экземпляров синтаксического анализатора, хранящихся в пуле. Используется для ускорения извлечения фактов. Рекомендуемое значение вычисляется по формуле: количество аппаратных потоков + 1. Значение по умолчанию **5**.

port. TCP-порт сервиса. Значение по умолчанию **9017**.

LOGGING – параметры ведения лог-файлов.

period. Периодичность создания новых лог-файлов. Возможные значения:

- **per_minute** – каждую минуту;
- **per_hour** – каждый час;
- **per_day** – каждый день;
- **midnight** – каждый день в полночь. Значение по умолчанию.

suffix. Формат названия лог-файла. Значение по умолчанию **%%Y%%m%%d**.

Object Detection Service

Сервис Object Detection Service настраивается в конфигурационном файле app.config.

Структура файла:

[DEFAULT]

debug = {Записывать в лог-файл отладочные сообщения}

port = {TCP-порт сервиса}

[LOGGING]

period = {Периодичность создания новых лог-файлов}

suffix = {Формат названия лог-файлов}

DEFAULT – основные настройки сервиса:

debug. Записывать в лог-файл отладочные сообщения. Возможные значения: **True**, **False**. Значение по умолчанию **True**.

port. TCP-порт сервиса. Значение по умолчанию **9047**.

LOGGING – параметры ведения лог-файлов.

period. Периодичность создания новых лог-файлов. Возможные значения:

- **per_minute** – каждую минуту;
- **per_hour** – каждый час;
- **per_day** – каждый день;
- **midnight** – каждый день в полночь. Значение по умолчанию.

suffix. Формат названия лог-файла. Значение по умолчанию **%%Y%%m%%d**.

Справочная информация

Ознакомьтесь со справочной информацией:

- список [фактов](#), которые умеют извлекать сервисы Arjo;
- сторонние [компоненты](#), которые входят в состав сервисов Arjo.

Список извлекаемых фактов

Тип документа	Извлекаемые факты
Входящее письмо	Тип документа Дата документа Номер документа Дата документа «В ответ на» Номер документа «В ответ на» Организация-корреспондент Организация-получатель Адресат Подписант письма Исполнитель Тема письма, начинающаяся с «О» или «Об» и расположенная вверху первой страницы письма Гриф доступа: «Конфиденциально», «Для служебного пользования», «Коммерческая тайна» Количество листов
Договор (в том числе на английском языке)	Документы на русском языке: Тип документа Номер договора Дата договора Сумма Контрагенты Количество листов Валюта Документы на английском языке: Тип документа Номер договора Дата договора Контрагент Количество листов
Дополнительное соглашение	Номер документа Дата документа Номер ведущего документа Дата ведущего документа Контрагенты

Тип документа	Извлекаемые факты
Счет-фактура	Тип документа Номер документа Дата документа Номер исправления Дата исправления Контрагенты Сумма документа Сумма НДС Валюта документа Количество листов Наименование товара Код единицы измерения Наименование, условное обозначение единицы измерения Количество(объем) Цена за единицу измерения Сумма НДС, по строке Стоимость с НДС, по строке
Корректировочный счет-фактура	Тип документа Номер документа Дата документа Номер исправления корректировки Дата исправления корректировки Номер корректируемого документа Дата корректируемого документа Контрагенты Валюта документа Количество листов
Универсальный передаточный документ	Тип документа Номер документа Дата документа Номер исправления Дата исправления Контрагенты Сумма документа Сумма НДС Валюта документа Статус (функция) УПД Количество листов Наименование товара Код единицы измерения Наименование, условное обозначение единицы измерения Количество(объем) Цена за единицу измерения Сумма НДС, по строке Стоимость с НДС, по строке

Тип документа	Извлекаемые факты
Универсальный корректировочный документ	Тип документа Номер документа Дата документа Номер исправления корректировки Дата исправления корректировки Номер корректируемого документа Дата корректируемого документа Контрагенты Валюта документа Статус (функция) УКД Количество листов
Товарная накладная (ТОРГ-12)	Тип документа Номер документа Дата документа Контрагенты Сумма документа Сумма НДС Наименование документа-основания Дата документа-основания Номер документа-основания Количество листов Валюта Наименование товара Наименование единицы измерения Код единицы измерения по ОКЕИ Количество (масса нетто) Цена, за единицу Сумма НДС, по строке Сумма с учетом НДС, по строке
Акт выполненных работ	Тип документа Номер документа Дата документа Контрагенты Сумма документа Наименование документа-основания Дата документа-основания Номер документа-основания Количество листов Валюта Наименование работ, услуг Количество Единица измерения Цена за единицу, без НДС Сумма НДС, по строке Сумма с НДС, по строке

Тип документа	Извлекаемые факты
Счет на оплату (в том числе на английском языке)	Документы на русском языке: Тип документа Номер документа Дата документа Контрагенты БИК контрагента Р/с контрагента Сумма документа Сумма НДС Количество листов Валюта Документы на английском языке: Тип документа Номер документа Дата документа Контрагенты Дата документа-основания Номер документа-основания Количество листов
Паспорт гражданина РФ	Серия и номер паспорта ФИО гражданина Пол гражданина Дата выдачи паспорта Дата рождения Место рождения Наименование организации, выдавшей паспорт Код подразделения организации, выдавшей паспорт
Свидетельство о постановке на налоговый учет	Наименование организации/ФИО физического лица ОГРН (для организаций) ИНН КПП (для организаций) Дата постановки на налоговый учет Серия и номер документа Наименование инспекции, выдавшей свидетельство
Авиабилет (электронный)	Дата отправления Дата прибытия Город отправления Город прибытия Стоимость билета (сумма и код валюты) ФИО персоны
Железнодорожный билет (электронный)	Дата отправления Дата прибытия Маршрут Стоимость билета (сумма и код валюты) ФИО персоны

Тип документа	Извлекаемые факты
По умолчанию (рекомендуется использовать для документов, которые не удалось классифицировать)	Дата документа Организационно-правовая форма Наименование организации/ФИО индивидуального предпринимателя ИНН организации/индивидуального предпринимателя КПП организации/индивидуального предпринимателя ОГРН юридического лица ОГРНИП физического лица Сумма документа Валюта документа ФИО персоны Номер телефона Адрес электронной почты

Сторонние компоненты

Компонент	Лицензия	Проект
Common.Logging	Лицензия: Apache 2.0 Лицензионное соглашение	Ссылка на проект
Keras	Лицензия: MIT Лицензионное соглашение	Ссылка на проект
Matplotlib	Лицензия: своя Лицензионное соглашение	Ссылка на проект
Microsoft .NET Core	Лицензия: MIT Лицензионное соглашение	Ссылка на проект
Newtonsoft.Json	Лицензия: MIT Лицензионное соглашение	Ссылка на проект
NLog	Лицензия: BSD 3 Clause Лицензионное соглашение	Ссылка на проект
Npgsql.EntityFrameworkCore.PostgreSQL	Лицензия: своя Лицензионное соглашение	Ссылка на проект
NumPy	Лицензия: своя Лицензионное соглашение	Ссылка на проект
OpenCV	Лицензия: BSD 3 Clause Лицензионное соглашение	Ссылка на проект
opencv-python	Лицензия: BSD 3 Clause Лицензионное соглашение	Ссылка на проект
OpenCvSharp3	Лицензия: BSD 3 Clause Лицензионное соглашение	Ссылка на проект
Pandas	Лицензия: BSD 3 Clause Лицензионное соглашение	Ссылка на проект
Psycopg	Лицензия: своя Лицензионное соглашение	Ссылка на проект
scikit-image	Лицензия: своя Лицензионное соглашение	Ссылка на проект

Компонент	Лицензия	Проект
Scikit-Learn	Лицензия: BSD Лицензионное соглашение	Ссылка на проект
Serilog.Extensions.Logging	Лицензия: Apache 2.0 Лицензионное соглашение	Ссылка на проект
TensorFlow	Лицензия: Apache 2.0 Лицензионное соглашение	Ссылка на проект
Tesseract	Лицензия: Apache 2.0 Лицензионное соглашение	Ссылка на проект
Tesseract .Net Wrapper	Лицензия: Apache 2.0 Лицензионное соглашение	Ссылка на проект
Tomita-parser	Лицензия: Mozilla Public License Version 2.0 Лицензионное соглашение	Ссылка на проект
Tornado	Лицензия: Apache 2.0 Лицензионное соглашение	Ссылка на проект