



Aio One

Directum Aio One 1.3. руководство
администратора Directum Aio (Linux).
Версия 2023.6

Содержание

Администрирование сервисов Ario (Linux)	3
Установка сервисов Ario.....	3
Подготовка к установке.....	3
Порядок установки.....	9
Удаление сервисов Ario.....	17
Обновление сервисов Ario.....	17
Особенности обновления моделей.....	18
Конфигурирование настроек Ario.....	19
Smart Service.....	29
Identity Service.....	31
Publishing Service.....	33
Text Extractor Service.....	34
Document Disassembler Service.....	38
Document Assembler Service.....	40
Text Classifier Service.....	42
Image Classifier Service.....	44
Fact Extractor Base Service.....	45
Fact Extractor Rules Service.....	47
Fact Extractor Learn Service.....	48
Object Detection Service.....	51
Storage Service.....	53
Elastic APM Service.....	53
Portainer.....	54
RabbitMQ.....	55
PostgreSQL.....	55
InfluxDB.....	56
Grafana.....	56
Telegraf.....	57
Общие параметры.....	57
Повторная активация сервисов.....	58
Справочная информация	59
Виды обрабатываемых документов.....	59
Список извлекаемых фактов.....	60
Ограничения при извлечении.....	68
Список распознаваемых языков.....	70

Администрирование сервисов Ario (Linux)

В главе содержатся описания [установки](#), [обновления](#) и [настройки](#) сервисов Ario в операционных системах на базе Linux.

Если срок действия лицензии закончился, [повторно активируйте сервисы](#).

Перед установкой и настройкой сервисов Ario ознакомьтесь с описанием их архитектуры и схемой обработки документов.

Установка сервисов Ario

Сервисы Ario рекомендуется устанавливать на отдельный от системы Directum RX сервер. Для этого:

1. Скачайте и установите необходимые для работы сервисов программные компоненты. Например, установите [Docker Engine](#).
2. [Выполните подготовительные действия](#).
3. [Установите сервисы](#) с помощью инструмента Directum Launcher.

Подготовка к установке

1. Откройте [необходимые порты](#) для работы сервисов и компонентов Directum Ario.
2. Для загрузки лог-файлов в Elasticsearch и их последующего анализа:
 - установите поисковую систему Elasticsearch и панель управления Kibana. Подробнее см. в инструкции по установке Directum RX, раздел «Установка поисковой системы Elasticsearch»;
 - [настройте шаблоны индекса в Kibana](#);
 - [установите и настройте сервис Filebeat](#).

Подробнее о работе инструментов см. в документации [Elasticsearch](#), [Kibana](#), [Filebeat](#).

2. Запросите в службе поддержки Directum ключ активации сервисов.
ПРИМЕЧАНИЕ. Файл активации имеет ограниченный период действия. Активацию необходимо выполнить в период, указанный службой поддержки.
3. Скопируйте полученный файл ActivationKey.aak в локальную папку. После установки сервисов активация выполняется автоматически. Копия файла с активационным ключом создается в папке с сервисами и при запуске Directum Smart Service заменяется на файл с лицензионным ключом LicenseKey.alk.

Список портов для подключения к сервисам Ario

Порт	Назначение
5432	Порт для подключения к базе данных. Если СУБД установлена локально, то значение по умолчанию берется из ветки реестра HKEY_LOCAL_MACHINE\SOFTWARE\PostgreSQL\Services\postgresql-x64-9.6\Port

Порт	Назначение
5672	Порт для подключения к RabbitMQ
15672	Дополнительный порт для подключения к панели управления RabbitMQ
9007	Порт для подключения к сервису Directum Text Extractor Service
9008	Порт для подключения к сервису Directum Image Classifier Service
9017	Порт для подключения к сервису Directum Fact Extractor Base Service
9018	Порт для подключения к сервису Directum Fact Extractor Rules Service
9019	Порт для подключения к сервису Directum Fact Extractor Learn Service
9021	Порт для подключения к сервису Directum Storage Service
9028	Порт для подключения к сервису Directum Document Disassembler Service
9029	Порт для подключения к сервису Directum Document Assembler Service
9030	Порт для подключения к сервису Directum Identity Service
9047	Порт для подключения к сервису Directum Object Detection Service
9057	Порт для подключения к сервису Directum Publishing Service
9082	Порт для подключения к сервису Directum Elastic APM Service
61100	Порт для подключения к сервису Directum Smart Service

Установка Docker Engine

Порядок установки компонентов Docker Engine зависит от используемого дистрибутива Linux.

В разделе приведен порядок установки на [Альт Сервер 10.1](#), [Альт Сервер 9.1](#), [Astra Linux Common Edition 2.12.40 \(Орел\)](#) и [РЕД ОС 7.3](#). Подробнее о порядке установки Docker на Ubuntu 20.04 LTS см. в документации Docker статью [Install Docker Engine](#).

Альт Сервер 10.1

Из-за специфических настроек безопасности в системе Альт Сервер 10.1 у пользователей по умолчанию отсутствуют права на выполнение команды `sudo`. Чтобы получить права, выполните соответствующие [настройки](#).

Рекомендуется всегда запускать сервис `docker` с помощью команды `sudo`. При необходимости можно настроить запуск сервиса `docker` без использования этой команды. Для этого нужно добавить текущего пользователя в группу `docker`. При этом следует учитывать, что добавление пользователей в группу `docker` повышает риск несанкционированного доступа к системным файлам.

1. Получите список обновлений и установите Docker Engine. Для этого выполните команды:

```
sudo apt-get update
sudo apt-get install docker-ce
```

2. Проверьте установленную версию Docker. Она должна быть не ниже 20. Для этого выполните команду:

```
sudo docker -v
```

3. Проверьте, что сервис docker запустился корректно:

```
sudo systemctl status docker
sudo docker ps
```

Если сервис установился и запустился успешно, первая команда возвращает статус **active**, вторая – пустой список контейнеров.

4. Запустите сервис вручную, если он не запустился автоматически. Для этого выдайте разрешение на автоматический запуск docker вместе с системой, последовательно выполнив команды:

```
sudo systemctl enable --now docker
sudo systemctl start docker
```

5. В конфигурационном файле /etc/docker/daemon.json в параметре **live-restore** установите значение **false**. Это необходимо для корректной работы компонентов Docker Engine.

6. Перезапустите сервис docker:

```
sudo systemctl restart docker
```

7. Проверьте, что сервис запустился корректно:

```
sudo systemctl status docker
```

Настройка прав для выполнения команды sudo

1. Переключитесь с текущего пользователя на пользователя root. Для этого выполните команду:

```
su -
```

2. Введите пароль пользователя **root**, заданный при установке операционной системы.

3. Откройте файл /etc/sudoers. Для этого выполните команду:

```
visudo
```

ПРИМЕЧАНИЕ. В результате указанный файл откроется в редакторе vi. Подробнее об использовании редактора см. в статье [Работа с редактором vi](#).

4. Раскомментируйте строку WHEEL_USERS ALL=(ALL) ALL, удалив символ #, и сохраните файл.

5. Проверьте, что пользователь добавлен в группу **wheel**. Пример:

```
cat /etc/group | grep wheel
```

При необходимости включите пользователя в группу:

```
usermod -aG wheel <имя пользователя>
```

6. Перезайдите в ОС под текущим пользователем.

Альт Сервер 9.1

Из-за специфических настроек безопасности в системе Альт Сервер 9.1 у пользователей по умолчанию отсутствуют права на выполнение команды `sudo`. Чтобы получить права, выполните соответствующие [настройки](#).

Рекомендуется всегда запускать сервис `docker` с помощью команды `sudo`. При необходимости можно настроить запуск сервиса `docker` без использования этой команды. Для этого нужно добавить текущего пользователя в группу `docker`. При этом следует учитывать, что добавление пользователей в группу `docker` повышает риск несанкционированного доступа к системным файлам.

1. Получите список обновлений. Для этого выполните команды:

```
sudo apt-get update
sudo apt-get install ca-certificates curl
```

2. Подключите репозиторий **Sisyphus**:

```
sudo apt-repo add sisyphus
```

3. Обновите информацию из репозитория:

```
sudo apt-get update
```

4. Убедитесь, что установка `Docker` будет производиться из репозитория **Sisyphus**, а не из репозитория по умолчанию. Для этого выполните команду:

```
sudo apt-cache policy docker-engine
```

Пример ответа:

```
установлен: (нет)
```

```
кандидат: 20.10.6-alt1:sisyphus+271055.100.1.1@1619787705
```

ПРИМЕЧАНИЕ. Если вместо ответа возвращается ошибка или версия предлагаемого к установке пакета ниже 20, обратитесь в службу поддержки.

5. Установите `Docker`:

```
sudo apt-get install docker-engine
```

ПРИМЕЧАНИЕ. Вместе с `Docker` обязательно должны установиться пакеты **containerd.io** и **docker-ce-cli**.

6. Проверьте установленную версию `Docker`. Она должна быть не ниже 20. Для этого выполните команду:

```
sudo docker -v
```

7. Проверьте, что сервис `docker` запустился корректно:

```
sudo systemctl status docker
sudo docker ps
```

Если сервис установился и запустился успешно, первая команда возвращает статус **active**; вторая – пустой список контейнеров.

8. Запустите сервис вручную, если он не запустился автоматически. Для этого выдайте разрешение на автоматический запуск `docker` вместе с системой, последовательно выполнив команды:

```
sudo systemctl enable --now docker
sudo systemctl start docker
```

9. В конфигурационном файле `/etc/docker/daemon.json` в параметре **live-restore** установите значение **false**. Это необходимо для корректной работы компонентов Docker Engine.
10. Перезапустите сервис `docker`:


```
sudo systemctl restart docker
```
11. Отключите репозиторий **Sisyphus**:


```
sudo apt-repo rm sisyphus
```
12. Проверьте, что сервис запустился корректно:


```
sudo systemctl status docker
```

Astra Linux Common Edition 2.12.40

Рекомендуется всегда запускать сервис `docker` с помощью команды `sudo`. При необходимости можно настроить запуск сервиса `docker` без использования этой команды. Для этого нужно добавить текущего пользователя в группу `docker`. При этом следует учитывать, что добавление пользователей в группу `docker` повышает риск несанкционированного доступа к системным файлам.

ВАЖНО. В инструкции приводятся готовые примеры кода и командных строк. При копировании текста убедитесь в отсутствии лишних пробелов, переносов строк и спецсимволов.

1. Получите список обновлений. Для этого выполните команды:


```
sudo apt update
sudo apt install apt-transport-https ca-certificates curl software-properties-common gnupg2
```
2. Добавьте ключ официального репозитория **Docker**:


```
sudo curl -fsSL https://download.docker.com/linux/debian/gpg | sudo apt-key add -
```
3. Добавьте репозиторий **Docker** в систему в зависимости от типа прав пользователя:
 - при работе под пользователем с правами **root** выполните команду:


```
echo "deb [arch=amd64] https://download.docker.com/linux/debian buster stable" >> /etc/apt/sources.list
```
 - при работе под другим пользователем выполните команду:


```
sudo bash -c "echo 'deb [arch=amd64] https://download.docker.com/linux/debian buster stable' >> /etc/apt/sources.list"
```
4. Обновите информацию из репозитория:


```
sudo apt update
```
5. Убедитесь, что установка `Docker` будет производиться из репозитория **Docker**, а не из репозитория по умолчанию. Для этого выполните команду:


```
sudo apt-cache policy docker-ce
```

Пример ответа:

```
docker-ce:
  установлен: (нет)
```

```
кандидат: 5:20.10.6~3-0~debian-buster
```

ПРИМЕЧАНИЕ. Если вместо ответа возвращается ошибка или версия предлагаемого к установке пакета ниже 20, обратитесь в службу поддержки.

- Установите Docker:

```
sudo apt install docker-ce
```

ПРИМЕЧАНИЕ. Вместе с Docker обязательно должны установиться пакеты **containerd.io** и **docker-ce-cli**.

- Проверьте, что сервис запустился:

```
sudo systemctl status docker
```

Команда должна вернуть статус «active», выделенный зеленым цветом.

- Включите сервис вручную, если он не запустился автоматически. Для этого выдайте разрешение на автоматический запуск docker вместе с системой, последовательно выполнив команды:

```
sudo systemctl enable --now docker
sudo systemctl start docker
```

РЕД ОС 7.3

Рекомендуется всегда запускать сервис docker с помощью команды sudo. При необходимости можно настроить запуск сервиса docker без использования этой команды. Для этого нужно добавить текущего пользователя в группу docker. При этом следует учитывать, что добавление пользователей в группу docker повышает риск несанкционированного доступа к системным файлам.

- Установите Docker Engine. Для этого выполните команду:

```
dnf install docker-ce
```

- Проверьте установленную версию Docker. Она должна быть не ниже 20. Для этого выполните команду:

```
sudo docker -v
```

- Проверьте, что сервис запустился корректно:

```
sudo systemctl status docker
sudo docker ps
```

Если сервис установился и запустился успешно, первая команда возвращает статус **active**; вторая – пустой список контейнеров.

- Запустите сервис вручную, если он не запустился автоматически. Для этого выдайте разрешение на автоматический запуск docker вместе с системой, последовательно выполнив команды:

```
sudo systemctl enable --now docker
sudo systemctl start docker
```

- В конфигурационном файле /etc/docker/daemon.json в параметре **live-restore** установите значение **false**. Это необходимо для корректной работы компонентов Docker Engine.

6. Перезапустите сервис docker:

```
sudo systemctl restart docker
```

7. Проверьте, что сервис запустился корректно:

```
sudo systemctl status docker
```

Подробнее см. в документации РЕД ОС статью [«Установка и настройка docker»](#).

Сервис Filebeat

Автоматическая загрузка лог-файлов в Elasticsearch выполняется с помощью сервиса Filebeat. Сервис входит в линейку продуктов Elasticsearch и распространяется бесплатно.

ВАЖНО. В инструкции приводятся готовые примеры кода и командных строк. При копировании текста убедитесь в отсутствии лишних пробелов и спецсимволов.

Чтобы настроить загрузку лог-файлов с помощью Filebeat:

1. Скачайте и установите Filebeat 7.14.0. Для этого выполните команды:

```
curl -L -O https://artifacts.elastic.co/downloads/beats/filebeat/filebeat-7.14.0-linux-x86_64.tar.gz
tar xzvf filebeat-7.14.0-linux-x86_64.tar.gz
```

2. Настройте сервис с помощью конфигурационного файла filebeat.yml. Путь до конфигурационного файла по умолчанию: /etc/filebeat/filebeat.yml.

- в секции **filebeat.inputs** укажите параметры получения и декодирования файлов:
 - paths** – пути до лог-файлов сервисов Ario в формате <путь до каталога с сервисами>/logs/<название сервиса>. Пути до каталогов с лог-файлами указываются в конфигурационном файле сервисов .env;

```
json.keys_under_root – true;
```

```
json.overwrite_keys – true;
```

```
json.add_error_key – true;
```

```
json.message_key – message;
```

- в секции **output.elasticsearch** укажите параметры загрузки данных в Elasticsearch:

```
hosts – адрес Elasticsearch;
```

- в остальных секциях оставьте значения по умолчанию.

3. Запустите Filebeat с помощью команды:

```
sudo ./filebeat
```

Порядок установки

Для установки сервисов Ario в операционной системе на базе Linux используются docker-контейнеры. Каждому сервису соответствует отдельный контейнер.

Перед установкой ознакомьтесь с типовыми требованиями и убедитесь, что ваше аппаратное и программное обеспечение подходит для установки сервисов. Требуемая конфигурация оборудования определяется с учетом объема документов, который обрабатывается сервисами Ario:

- если объем документопотока в месяц составляет до 90 000 документов, все сервисы Directum Ario можно самостоятельно установить на один сервер;
- если объем документопотока в месяц превышает 90 000 документов, рекомендуется установить Directum Text Extractor Service и другие сервисы на отдельные виртуальные или физические серверы. Чтобы выполнить распределенную установку, обратитесь в службу поддержки Directum.

ПРИМЕЧАНИЕ. Объем документопотока рассчитан исходя из средних значений: объем одного документа – 2 страницы, время обработки – 8 часов в день, 22 дня в месяц.

Чтобы установить сервисы Directum Ario на один сервер:

1. Архив с Directum Launcher распакуйте в локальную папку на сервере с помощью команды:

```
tar -xvf <Имя архива> -C <Имя папки>
```

ВАЖНО. Для корректной установки общий путь к файлам должен быть не более 256 символов. Также он не должен содержать пробелы, символы кириллицы, запятые и спецсимволы. Поэтому используйте, например, папку /srv/DirectumLauncher. В зависимости от настроек операционной системы для дальнейших действий могут потребоваться права суперпользователя.

2. В корень папки с Directum Launcher скопируйте архив с сервисами Ario.tar.gz. Если на сервере отсутствует доступ к сети Интернет, также скопируйте архив с базовыми образами для сервисов ArioBaselimages.tar.gz.
3. Добавьте компоненты в Directum Launcher. Для этого выполните команду:

```
./do.sh components add_all
```

4. Создайте конфигурационный файл config.yml на основе файла DirectumLauncher/etc/config.yml.example. Сгенерируйте настройки установки сервисов с помощью команды::

```
./do.sh ario generate_config_yaml
```

ПРИМЕЧАНИЕ. Если конфигурационный файл config.yml был сгенерирован ранее, то создавать новый не нужно. По команде в текущем конфигурационном файле добавляется секция с параметрами установки сервисов Directum Ario.

5. Откройте конфигурационный файл config.yml и в секции **variables** задайте:
 - параметры подключения к базам данных Smart Service или Identity Service. Для этого в строках подключения к базам данных **ARIO_PG_CONNECTION_STRING** и **ARIO_PG_CONNECTION_STRING_DIS** соответственно задайте параметры:

User ID – логин для подключения к базе данных. Значение по умолчанию **postgres**;

Password – пароль для подключения к базе данных. Использование пустого пароля не допускается;

Host – имя компьютера, на котором установлена СУБД. Если при установке PostgreSQL имя компьютера по умолчанию было изменено, укажите имя сервера в формате <Имя компьютера>/<Имя экземпляра>;

Port – порт для подключения к базе данных. Значение по умолчанию **5432**;

Datadase – имя базы данных. Укажите имя ранее созданной базы данных или задайте имя новой. Если указать имя новой базы данных, то она создается при первом запуске сервисов. Значения по умолчанию: **Ario_Smart_Service** – для базы данных Smart Service; **Ario_Identity_Service** – для базы данных Identity Service.

- значения остальных параметров:

ARIO_STORAGE_SERVICE_URL – хост, на котором необходимо развернуть Directum Storage Service. Значение по умолчанию **localhost**.

ARIO_STORAGE_SERVICE_PORT – порт для подключения к сервису Directum Storage Service. Значение по умолчанию **9021**;

ARIO_IDENTITY_SERVICE_PORT – порт для подключения к сервису Directum Identity Service. Значение по умолчанию **9030**;

ARIO_LOG_PATH – относительный или полный путь до папки с лог-файлами и файлами трассировки. Значение по умолчанию **/opt/directum/ario/logs**;

ARIO_INSTALL_PATH – относительный или полный путь до папки установки сервисов Directum Ario. Значение по умолчанию **/opt/directum/ario**;

ARIO_ACTIVATION_KEY_PATH – относительный или полный путь до ключа активации.

Пример секции **variables**:

`variables:`

```
ARIO_PG_CONNECTION_STRING: {строка подключения к базе данных Smart Service};
ARIO_PG_CONNECTION_STRING_DIS: {строка подключения к базе данных Smart Service};
ARIO_QUEUE_CONNECTION_STRING: {строка подключения к очереди сообщений RabbitMQ}
ARIO_STORAGE_SERVICE_URL: {хост для хранения оперативных данных}
ARIO_STORAGE_SERVICE_PORT: {порт подключения к сервису Storage Service}
ARIO_IDENTITY_SERVICE_PORT: {порт подключения к сервису Identity Service}
ARIO_LOG_PATH: {путь до лог-файлов}
ARIO_INSTALL_PATH: {путь до папки установки}
ARIO_ACTIVATION_KEY_PATH: {путь до ключа активации}
```

6. В конфигурационном файле `config.yml` в секции **services_config** проверьте значения параметров установки сервисов.

Структура секции **services_config**:

`services_config:`

`ArioSmartService:`

```
PORT: {порт подключения к сервису Smart Service}
CONNECTION_STRING: "{{ ARIO_PG_CONNECTION_STRING }}"
STORAGE_ENDPOINT: "{{ ARIO_STORAGE_SERVICE_URL }}:{{ ARIO_STORAGE_SERVICE_PORT }}"
IDENTITY_SERVICE: http://localhost:{{ ARIO_IDENTITY_SERVICE_PORT }}
QUEUE_CONNECTION_STRING: "{{ ARIO_QUEUE_CONNECTION_STRING }}"
```

`ArioStorageService:`

```
PORT: "{{ ARIO_STORAGE_SERVICE_PORT }}"
```

`ArioElasticApm:`

```
PORT: {порт подключения к сервису Elastic APM Service}
TRACES_PATH: ./traces
```

`ArioIdentityService:`

```
PORT: "{{ ARIO_IDENTITY_SERVICE_PORT }}"
CONNECTION_STRING: "{{ ARIO_PG_CONNECTION_STRING_DIS }}"
```

`ArioPublishingService:`

```
PORT: {порт подключения к сервису Publishing Service}
QUEUE_CONNECTION_STRING: "{{ ARIO_QUEUE_CONNECTION_STRING }}"
```

`ArioTextExtractor:`

```
PORT: {порт подключения к сервису Text Extractor Service}
QUEUE_CONNECTION_STRING: "{{ ARIO_QUEUE_CONNECTION_STRING }}"
```

`ArioDocumentDisassemblerService:`

```
PORT: {порт подключения к сервису Document Disassembler Service}
QUEUE_CONNECTION_STRING: "{{ ARIO_QUEUE_CONNECTION_STRING }}"
```

```

ArioDocumentAssemblerService:
  PORT: {порт подключения к сервису Document Assembler Service}
  QUEUE_CONNECTION_STRING: "{{ ARIO_QUEUE_CONNECTION_STRING }}"
ArioFactExtractorService:
  PORT: {порт подключения к сервису Fact Extractor Base Service}
  QUEUE_CONNECTION_STRING: "{{ ARIO_QUEUE_CONNECTION_STRING }}"
ArioFactExtractorRulesService:
  PORT: {порт подключения к сервису Fact Extractor Rules Service}
  QUEUE_CONNECTION_STRING: "{{ ARIO_QUEUE_CONNECTION_STRING }}?heartbeat=300"
ArioFactExtractorLearnService:
  PORT: {порт подключения к сервису Fact Extractor Learn Service}
  CONNECTION_STRING: "{{ ARIO_PG_CONNECTION_STRING }}"
  QUEUE_CONNECTION_STRING: "{{ ARIO_QUEUE_CONNECTION_STRING }}?heartbeat=36000"
ArioObjectDetectionService:
  PORT: {порт подключения к сервису Object Detection Service}
  QUEUE_CONNECTION_STRING: "{{ ARIO_QUEUE_CONNECTION_STRING }}"
ArioImageClassifierService:
  PORT: {порт подключения к сервису Image Classifier Service}
  QUEUE_CONNECTION_STRING: "{{ ARIO_QUEUE_CONNECTION_STRING }}"
ArioTextClassifierService:
  PORT: {порт подключения к сервису Text Classifier Service}
  CONNECTION_STRING: "{{ ARIO_PG_CONNECTION_STRING }}"
  QUEUE_CONNECTION_STRING: "{{ ARIO_QUEUE_CONNECTION_STRING }}"

```

где:

ArioSmartService – параметры установки сервиса Directum Smart Service:

- **PORT** – порт для подключения к сервису. Значение по умолчанию **61100**;
- **CONNECTION_STRING** – строка подключения к базе данных Smart Service. Подставляется значение, указанное для параметра **ARIO_PG_CONNECTION_STRING** в секции **variables**;
- **STORAGE_ENDPOINT** – строка подключения к сервису хранения объектов Storage Service. Указывается без протокола http/https. Пример значения: **localhost:9000**;
- **IDENTITY_SERVICE** – адрес сервиса Identity Service. Рекомендуется не менять значение по умолчанию **localhost**;
- **QUEUE_CONNECTION_STRING** – строка подключения к RabbitMQ. Подставляется значение, указанное для параметра **ARIO_QUEUE_CONNECTION_STRING** в секции **variables**.

ArioStorageService – параметры установки сервиса Directum Storage Service:

- **PORT** – порт для подключения к сервису. Значение по умолчанию **9021**;

ArioElasticApm – параметры установки сервиса Directum Elastic APM Service:

- **PORT** – порт для подключения к сервису. Значение по умолчанию **9082**;
- **TRACES_PATH** – папка, в которую будут записываться файлы трассировки запросов к сервисам Ario. Значение по умолчанию **<Путь до лог-файлов>\Traces**;

ArioIdentityService – параметры установки сервиса Directum Identity Service:

- **PORT** – порт для подключения к сервису. Значение по умолчанию **9030**;
- **CONNECTION_STRING** – строка подключения к базе данных Smart Service. Подставляется значение, указанное для параметра **ARIO_PG_CONNECTION_STRING** в секции **variables**;

ArioPublishingService – параметры установки сервиса Directum Publishing Service:

- **PORT** – порт для подключения к сервису. Значение по умолчанию **9057**;

- **QUEUE_CONNECTION_STRING** – строка подключения к RabbitMQ. Подставляется значение, указанное для параметра **ARIO_QUEUE_CONNECTION_STRING** в секции **variables**;

ArioTextExtractorService – параметры установки сервиса Directum Text Extractor Service:

- **PORT** – порт для подключения к сервису. Значение по умолчанию **9007**;
- **QUEUE_CONNECTION_STRING** – строка подключения к RabbitMQ. Подставляется значение, указанное для параметра **ARIO_QUEUE_CONNECTION_STRING** в секции **variables**;

ArioDocumentDisassemblerService – параметры установки сервиса Directum Document Disassembler Service:

- **PORT** – порт для подключения к сервису. Значение по умолчанию **9028**;
- **QUEUE_CONNECTION_STRING** – строка подключения к RabbitMQ. Подставляется значение, указанное для параметра **ARIO_QUEUE_CONNECTION_STRING** в секции **variables**;

ArioDocumentAssemblerService – параметры установки сервиса Directum Document Assembler Service:

- **PORT** – порт для подключения к сервису. Значение по умолчанию **9029**;
- **QUEUE_CONNECTION_STRING** – строка подключения к RabbitMQ. Подставляется значение, указанное для параметра **ARIO_QUEUE_CONNECTION_STRING** в секции **variables**;

ArioFactExtractorService – параметры установки сервиса Directum Fact Extractor Base Service:

- **PORT** – порт для подключения к сервису. Значение по умолчанию **9017**;
- **QUEUE_CONNECTION_STRING** – строка подключения к RabbitMQ. Подставляется значение, указанное для параметра **ARIO_QUEUE_CONNECTION_STRING** в секции **variables**;

ArioFactExtractorRulesService – параметры установки сервиса Directum Fact Extractor Rules Service:

- **PORT** – порт для подключения к сервису. Значение по умолчанию **9018**;
- **QUEUE_CONNECTION_STRING** – строка подключения к RabbitMQ. Подставляется значение, указанное для параметра **ARIO_QUEUE_CONNECTION_STRING** в секции **variables**;

ArioFactExtractorLearnService – параметры установки сервиса Directum Fact Extractor Learn Service:

- **PORT** – порт для подключения к сервису. Значение по умолчанию **9019**;
- **QUEUE_CONNECTION_STRING** – строка подключения к RabbitMQ. Подставляется значение, указанное для параметра **ARIO_QUEUE_CONNECTION_STRING** в секции **variables**;
- **CONNECTION_STRING** – строка подключения к базе данных Smart Service. Подставляется значение, указанное для параметра **ARIO_PG_CONNECTION_STRING** в секции **variables**;

ArioObjectDetectionService – параметры установки сервиса Directum Object Detection Service:

- **PORT** – порт для подключения к сервису. Значение по умолчанию **9047**;
- **QUEUE_CONNECTION_STRING** – строка подключения к RabbitMQ. Подставляется значение, указанное для параметра **ARIO_QUEUE_CONNECTION_STRING** в секции **variables**;

ArioImageClassifierService – параметры установки сервиса Directum Image Classifier Service:

- **PORT** – порт для подключения к сервису. Значение по умолчанию **9008**;
- **QUEUE_CONNECTION_STRING** – строка подключения к RabbitMQ. Подставляется значение, указанное для параметра **ARIO_QUEUE_CONNECTION_STRING** в секции **variables**;

ArioTextClassifierService – параметры установки сервиса Directum Text Classifier Service:

- **PORT** – порт для подключения к сервису. Значение по умолчанию **9007**;
- **QUEUE_CONNECTION_STRING** – строка подключения к RabbitMQ. Подставляется значение, указанное для параметра **ARIO_QUEUE_CONNECTION_STRING** в секции **variables**;
- **CONNECTION_STRING** – строка подключения к базе данных Smart Service. Подставляется значение, указанное для параметра **ARIO_PG_CONNECTION_STRING** в секции **variables**.

При необходимости скорректируйте значения.

7. Установите сервисы Directum Ario. Для этого выполните команду:

```
./do.sh ario install
```

8. Дождитесь окончания установки.
9. Добавьте рассчитанные лимиты объема оперативной памяти для Directum Text Extractor Service и Directum Fact Extractor Learn Service в конфигурационный файл `deploy.yml`. Для этого измените значение параметра **memory** в соответствующих секциях файла, указав рассчитанный объем в байтах.

ПРИМЕЧАНИЕ. Чтобы в параметре **memory** указать значение в мегабайтах или гигабайтах, после числа введите **m** или **g** соответственно. Например, для лимита оперативной памяти 16384 мегабайта будет соответствовать значение **16384m**.

Настройки сервиса Directum Text Extractor Service соответствует секция **dtes**, сервису Directum Fact Extractor Learn Service – **dfesl**:

```

deploy:
  mode: replicated
  replicas: ${dtes_replicas}
  placement:
    constraints:
      - "node.labels.service == ${dtes_node_label}"
    max_replicas_per_node: 1
  restart_policy:
    condition: any
    delay: 15s
    window: 180s
  resources:
    limits:
      memory: 16384m

```

ПРИМЕЧАНИЕ. Если для распознавания и обучения используется два экземпляра сервиса Directum Fact Extractor Learn Service, то в первом случае лимит оперативной памяти имеет постоянное значение – 2 ГБ. Лимит оперативной памяти для обучения определяется с учетом количества обрабатываемых документов.

Если для распознавания и обучения используется один экземпляр Directum Fact Extractor Learn Service, лимит оперативной памяти для распознавания также будет равен 2 ГБ, а итоговый лимит памяти равен объему памяти для обучения согласно таблице конфигурации оборудования для сервиса Directum Fact Extractor Learn Service. Например, при 2 ГБ для распознавания и 16 ГБ для обучения на 400 документах итоговый лимит 16 ГБ. Таким образом, итоговый лимит равен наибольшему значению лимита для экземпляра сервиса.

10. Убедитесь, что в параметрах конфигурационного файла `.env` указаны значения:

- **timezone** – текущий часовой пояс;
- **dss_allow_local_path** – **true**;
- **dtcs_allow_local_path** – **True**.

11. Если дополнительно необходимо изменить другие настройки определенного сервиса, внесите корректировки в [конфигурационный файл .env](#) в соответствующую секцию. Например, чтобы изменить время хранения ссылок на объекты, внесите изменения в секцию **SmartService** в параметр **dss_storage_uri_lifetime_seconds**.

12. После изменения настроек перезапустите сервисы. Для этого сначала остановите работу сервисов:

```
./do.sh ario down
```

Затем повторно запустите сервисы:

```
./do.sh ario up
```

В результате сервисы запустятся с учетом обновленных настроек в конфигурационном файле.

В результате установки:

- из архива загружаются docker-образы;
- запускаются контейнеры с сервисами Directum Ario;
- автоматически импортируются базовые *модели извлечения фактов* и базовые *модели классификации* документов.

СОВЕТ. Если в результате установки произошла ошибка импорта моделей, их можно установить [вручную](#) с помощью скриптов.

Ручной импорт базовых моделей извлечения фактов и классификации

1. Импортируйте базовые *модели извлечения фактов*, если факты должны извлекаться на основе обучаемых моделей. Для этого запустите скрипт `import_models.py` из каталога `./etc/factextractor-learn-service/trained_models` с обязательным для заполнения параметром **-url**, в котором задается адрес до сервиса Directum Smart Service.

ПРИМЕЧАНИЕ. Если изменялись логин и пароль администратора сервисов Ario для аутентификации, заданные по умолчанию, то укажите их актуальные значения в параметрах **-login** и **-password**.

Пример команды:

```
python import_models.py -url http://smart:61100
```

СОВЕТ. В зависимости от настроек операционной системы может потребоваться указать конкретную версию Python, например `python3.7`, а вместо имени сервера – его IP-адрес.

В результате работы скрипта в Ario инициализируются базовые модели извлечения фактов.

2. Импортируйте базовые *модели классификации* документов.

Для первичной загрузки моделей в новый классификатор запустите скрипт `import_models.py` из каталога `./etc/textclassifier-service/trained_models` с обязательным для заполнения параметром **-url**, в котором задается адрес до сервиса Directum Smart Service.

ПРИМЕЧАНИЕ. Если изменялись логин и пароль администратора сервисов Ario для аутентификации, заданные по умолчанию, то укажите их актуальные значения в параметрах **-login** и **-password**.

Пример команды:

```
python import_models.py -url http://smart:61100
```

СОВЕТ. В зависимости от настроек операционной системы может потребоваться указать конкретную версию Python, например `python3.7`, а вместо имени сервера – его IP-адрес.

В результате создаются классификаторы по виду документа и по первым страницам, в которые импортируются соответствующие базовые модели классификации.

ПРИМЕЧАНИЕ. Если уже существуют классификаторы с типами **Kind** и **FirstPage**, то в результате работы скрипта обновятся их модели.

Удаление сервисов Ario

Порядок удаления сервисов Directum Ario зависит от используемой версии:

- [версии 2023.2 и выше](#);
- [версии ниже 2023.2](#).

Удаление сервисов версии 2023.2 и выше

1. Создайте резервную копию конфигурационного файла Directum RX. Для этого скопируйте файл config.yml на локальный компьютер.
2. Выполните команду:

```
./do.sh ario uninstall
```

В результате все папки с сервисами удалятся, кроме папки с лог-файлами и файлами трассировки. При необходимости удалите ее отдельно.

Удаление сервисов версии ниже 2023.2

Чтобы удалить сервисы Ario с сервера:

1. Сохраните текущий конфигурационный файл .env.
2. Остановите сервисы. Для этого выполните команду:


```
./stop-services.sh
```
3. Проверьте, что все сервисы остановлены, с помощью команды:


```
docker ps a
```
4. Удалите Docker-образы сервисов средствами Docker. Подробнее см. в документации Docker статью [docker rmi](#).
5. Удалите корневой каталог, в котором находятся скрипты управления сервисами.
6. При необходимости удалите хранилища данных (volumes) Portainer, Grafana, InfluxDB, Postgres, RabbitMQ и Minio средствами Docker. Например, если не планируете использовать текущие хранилища данных с новой версией сервисов. Подробнее см. в документации Docker статью [docker volume rm](#).

ВАЖНО. Данные из хранилищ удаляются без возможности восстановления.

Обновление сервисов Ario

Перед обновлением ознакомьтесь с типовыми требованиями и убедитесь, что ваше аппаратное и программное обеспечение актуально.

Чтобы обновить сервисы Directum Ario на новую версию:

1. Ознакомьтесь с особенностями обновления моделей.
2. Удалите старую версию сервисов и компонентов Directum Ario.
3. Установите новую версию.
4. Вручную перенесите настройки старой версии в новую.

В результате обновляются модели базовых грамматик и классификаторов. Если последние опубликованные в грамматиках и классификаторах модели пользовательские, то при обновлении сервисов они не меняются.

В версиях Ario различается механизм извлечения текстового слоя, поэтому после обновления сервисов обновите текстовый слой, а пользовательские модели:

- обучите заново, если до обновления они обучались «с нуля»;
- дообучите, если до обновления они только дообучались.

Это позволяет сохранить высокое качество извлечения фактов. Подробнее о порядке обновления текстового слоя см. в инструкции AVIA, раздел «Обновление текстового слоя».

ВАЖНО. В версии сервисов Ario 2022.12 добавлены базовые классификаторы с типами **Kind** и **FirstPage**. С этой версии регулярно улучшаться и обновляться будут только эти базовые классификаторы. Чтобы вместо предыдущих использовать актуальные модели классификации, в запросах передавайте идентификатор этих классификаторов.

Особенности обновления моделей

При обновлении сервисов **до версии 2022.7:**

- если в базе данных уже существуют классификаторы с именами **Базовый классификатор по виду документа** и **Базовый классификатор первая/не первая страница**, то в результате обновления для них создаются новые модели классификации;
- если в базе данных уже существуют модели извлечения фактов для базовых грамматик, то в результате установки создаются новые модели извлечения фактов.

При обновлении сервисов **до версии 2022.12:**

- создаются новые базовые классификаторы с типами **Kind** и **FirstPage**, в них загружаются новые модели классификации;
- если в базе данных уже существуют модели извлечения фактов для базовых грамматик, то в результате установки создаются новые модели извлечения фактов.

При обновлении сервисов, начиная **с версии 2022.12:**

- если в базе данных уже существуют классификаторы с типами **Kind** и **FirstPage**, то в результате обновления для них создаются новые модели классификации;
- если в базе данных уже существуют модели извлечения фактов для базовых грамматик, то в результате установки создаются новые модели извлечения фактов.

Все предыдущие модели классификации и извлечения фактов сохраняются в базе данных. При необходимости их можно [удалить с помощью Postman](#). Для этого выберите в коллекции DEL-запрос на удаление соответствующей модели, укажите **Id** модели и выполните запрос.

ВАЖНО. После обновления сервисов на версию 2022.12 в запросах на классификацию документов необходимо указывать идентификаторы новых базовых классификаторов. Чтобы узнать их, выполните запрос на просмотр списка классификаторов. В полученном ответе найдите классификаторы, для которых в параметре **type** указаны значения **Kind** (базовый классификатор по виду документа) и **FirstPage** (базовый классификатор по первой странице). Идентификатор классификаторов будет указан в параметре **classifierId**.

При обновлении сервисов, начиная **с версии 2023.1**:

- изменилась структура моделей извлечения фактов с многофактовой на однофактовую;
- в базовых моделях структура меняется автоматически. Чтобы перевести на однофактовую структуру пользовательские модели, которые обучены до версии 2023.1, переобучите их. Это упростит их дальнейшее дообучение.

Так как в версиях Ario различается механизм извлечения текстового слоя, после обновления сервисов до любой версии рекомендуется обновить текстовый слой, а также заложить дополнительное время на обучение или дообучение пользовательских моделей. Это позволяет сохранить высокое качество извлечения фактов. Подробнее о порядке обновления текстового слоя см. в инструкции AVIA, раздел «Обновление текстового слоя».

Конфигурирование настроек Ario

В процессе работы администратор может настраивать сервисы Ario, редактируя их конфигурационные файлы.

Чтобы настроить работу сервисов:

1. Остановите работу всех сервисов. Для этого выполните команду:

```
./do.sh ario down
```

2. Для внесения точечных изменений в работу сервисов скорректируйте значения параметров в соответствующих секциях конфигурационного файла `.env`. Файл находится в корневом каталоге сервисов:

- [Smart Service](#)
- [Identity Service](#)
- [Publishing Service](#)
- [Text Extractor Service](#)
- [Document Disassembler Service](#)
- [Document Assembler Service](#)
- [Text Classifier Service](#)
- [Image Classifier Service](#)
- [Fact Extractor Base Service](#)
- [Fact Extractor Rules Service](#)
- [Fact Extractor Learn Service](#)
- [Object Detection Service](#)
- [Storage Service](#)
- [Elastic APM Service](#)
- [Portainer](#)
- [RabbitMQ](#)
- [PostgreSQL](#)
- [InfluxDB](#)
- [Grafana](#)
- [Telegraf](#)

Кроме того, в конфигурационном файле есть [общие параметры](#), которые используются во всех сервисах Directum Ario.

3. По завершении настройки вновь запустите сервисы. Для этого выполните команду:

```
./do.sh ario up
```

При конфигурировании настроек для сервисов на Linux учитывайте следующие рекомендации и особенности:

1. Для удаленных серверов вместо имен рекомендуется указывать IP-адреса.
2. В именах и паролях пользователей не используйте символы # @ \ / : * ? " < > |.
3. Для названия сервисов используется аббревиация. Например, сервис Directum Text Classifier Service сокращен до dtcs, а Directum Publishing Service до dps и т.д.
4. Названия параметров для разных сервисов именуются по разным моделям:
 - для сервисов Directum Smart Service, Directum Identity Service, Directum Publishing Service, Directum Text Extractor Service, Directum Document Disassembler Service, Directum Document Assembler Service по модели: <аббревиатура сервиса>_<секция конфигурационного файла>_<параметр>. Например, **dss_serilog_elastic_apm**;
 - для остальных сервисов по модели: <аббревиатура сервиса>_<параметр>. Например, **dfesl_debug**.

Структура конфигурационного файла:

```
ario_version = {Версия сервисов Ario}
```

```
queues_message_ttl = {Время хранения сообщений в RabbitMQ в секундах}
```

```
# ContainersTimezone
```

```
timezone = {Часовой пояс внутри сервисов}
```

```
# Logs
```

```
log_path = {Корневой каталог для лог-файлов}
```

```
log_level = {Минимальный уровень протоколирования}
```

```
log_path_format = {Формат имени лог-файла}
```

```
dss_logs = {Подкаталог для лог-файлов Smart Service}
```

```
dtes_logs = {Подкаталог для лог-файлов Text Extractor Service}
```

```
dis_logs = {Подкаталог для лог-файлов Identity Service}
```

```
dps_logs = {Подкаталог для лог-файлов Publishing Service}
```

```
dfes_logs = {Подкаталог для лог-файлов Fact Extractor Service}
```

```
dfesr_logs = {Подкаталог для лог-файлов Fact Extractor Rules Service}
```

```
dfesl_logs = {Подкаталог для лог-файлов Fact Extractor Learn Service}
```

```
dods_logs = {Подкаталог для лог-файлов Object Detection Service}
```

```
dics_logs = {Подкаталог для лог-файлов Image Classifier Service}
```

```
dtcs_logs = {Подкаталог для лог-файлов Text Classifier Service}
```

```
# ServicePlacement
```

```
manager_node_label = {Label для управляющего сервера}
```

```
pg_node_label = {Label для сервера с СУБД PostgreSQL}
```

```
rabbit_node_label = {Label для сервера с брокером сообщений RabbitMQ}
```

```
storage_service_node_label = {Label для сервера с сервисом хранилищ}
```

```
dtes_node_label = {Label для сервера с Text Extractor Service}
```

```
dfes_node_label = {Label для сервера с Fact Extractor Service}
```

```
dfesr_node_label = {Label для сервера с Fact Extractor Rules Service}
```

```

dfesl_node_label = {Label для сервера с Fact Extractor Learn Service}
dods_node_label = {Label для сервера с Object Detection Service}
dics_node_label = {Label для сервера с Image Classifier Service}
dtcs_node_label = {Label для сервера с Text Classifier Service}

# Portainer
portainer_external_port = {TCP-порт Portainer}

# RabbitMQ
rabbit_replicas = {Количество запускаемых копий брокера сообщений}
rabbit_external_port = {TCP-порт брокера сообщений}
rabbit_user = {Имя пользователя брокера сообщений}
rabbit_password = {Пароль пользователя брокера сообщений}
rabbit_url = amqp://${rabbit_user}:${rabbit_password}@{Виртуальный хост}:{Порт}

# PostgreSQL
pg_replicas = {Количество запускаемых копий СУБД PostgreSQL}
pg_external_port = {TCP-порт СУБД PostgreSQL}
pg_host = {Адрес СУБД PostgreSQL}
pg_user = {Имя пользователя для подключения к СУБД PostgreSQL}
pg_password = {Пароль для подключения к СУБД PostgreSQL}
pg_database = {Имя базы данных PostgreSQL для Smart Service}
pg_database_dis = {Имя базы данных PostgreSQL для Identity Service}
pg_port = {Порт для подключения к СУБД PostgreSQL}
pg_connection = = User
ID=${pg_user};Password=${pg_password};Host=${pg_host};Port=${pg_port};Database=${p
g_database};Pooling={Использовать пул подключений к БД}

# StorageService
storage_service_replicas = {Количество запускаемых копий сервиса хранения
объектов}
storage_service_external_port = {TCP-порт сервиса хранения объектов}
storage_service_access_key = {Имя пользователя сервиса хранения объектов}
storage_service_secret_key = {Пароль пользователя сервиса хранения объектов}

# ElasticApm
elastic_apm_replicas = {Количество запускаемых копий сервиса Elastic APM Service}
elastic_apm_external_port = {TCP-порт сервиса Elastic APM Service}
elastic_apm_server_urls = {Адрес сервиса Elastic APM Service}
elastic_apm_log_level = {Минимальный уровень протоколирования сервиса Elastic APM
Service}
elastic_apm_transaction_sample_rate = {Регулярность отправки трейсов в сервис
Elastic APM Service}
elastic_apm_span_frames_min_duration = {Время, в течение которого в лог файл
записывается подробная информация об ошибке обработки документа. Указывается в
миллисекундах}
elastic_apm_metric_interval = {Интервал сбора метрик в секундах}
elastic_apm_capture_body = {Сообщения, которые передаются в сервис Elastic APM
Service}
elastic_apm_traces_path = {Путь до каталога с файлами трассировок}
elastic_apm_traces_file_name = {Базовое имя файлов трассировки}
elastic_apm_output_file_rotate_every_kb = {Максимальный размер файла трассировки}
elastic_apm_output_file_number_of_files = {Максимальное количество файлов
трассировки}

# InfluxDB

```

```
influxdb_replicas = {Количество запускаемых копий InfluxDB}
influxdb_user = {Имя пользователя InfluxDB}
influxdb_password = {Пароль пользователя InfluxDB}
```

[Grafana](#)

```
grafana_replicas = {Количество запускаемых копий Grafana}
grafana_external_port = {TCP-порт Grafana}
grafana_user = {Имя пользователя Grafana}
grafana_password = {Пароль пользователя Grafana}
```

[Telegraf](#)

```
telegraf_replicas = {Количество запускаемых копий Telegraf}
```

[SmartService](#)

```
dss_replicas = {Количество запускаемых копий сервиса Smart Service}
dss_address = {Адрес сервиса Smart Service}
dss_external_port = {TCP-порт сервиса Smart Service}
dss_connection_strings_default_connection = ${pg_connection}
dss_serilog_default = ${log_level}
dss_serilog_entity_framework_core = Error
dss_serilog_system = Error
dss_serilog_elastic_apm = Error
dss_serilog_application = SmartService
dss_serilog_name = RollingFile
dss_serilog_formatter = SmartService.Logger.SmartEcsTextFormatter, SmartService
dss_serilog_path_format = {Путь до каталога для логов Smart Service внутри контейнера}
dss_identity_service = {Адрес сервиса Identity Service}
dss_allow_local_path = {Разрешить обучение классификаторов из локального каталога}
dss_allow_anonymous_access = {Разрешить анонимную работу}
dss_anonymous_client_id = {Логин анонимного клиента}
dss_anonymous_client_secret = {Пароль анонимного клиента}
dss_storage_endpoint = {Адрес сервиса хранения объектов}
dss_storage_access_key = ${storage_service_access_key}
dss_storage_secret_key = ${storage_service_secret_key}
dss_storage_bucket = {Имя контейнера}
dss_storage_uri_lifetime_seconds = {Время хранения ссылок на объекты в секундах}
dss_storage_file_lifetime_hours = {Время хранения объектов в часах}
dss_storage_retry_connection_count = {Количество повторных попыток подключения к Storage Service}
dss_elastic_apm_server_urls = ${elastic_apm_server_urls}
dss_elastic_apm_transaction_sample_rate = ${elastic_apm_transaction_sample_rate}
dss_elastic_apm_span_frames_min_duration = ${elastic_apm_span_frames_min_duration}
dss_elastic_apm_metrics_interval = ${elastic_apm_metric_interval}
dss_elastic_apm_capture_body = ${elastic_apm_capture_body}
dss_elastic_apm_capture_body_content_types = "*"
dss_queues_message_ttl = ${queues_message_ttl}
dss_queue_connection = ${rabbit_url}
dss_queue_retry_connection_count = {Количество повторных попыток подключения к RabbitMQ}
```

[IdentityService](#)

```
dis_replicas = {Количество запускаемых копий сервиса Identity Service}
dis_external_port = {TCP-порт сервиса Identity Service}
dis_serilog_default = ${log_level}
```

```

dis_serilog_entity_framework_core = Error
dis_serilog_system = Error
dis_serilog_elastic_apm = Error
dis_serilog_application = IdentityService
dis_serilog_name = RollingFile
dis_serilog_formatter = IdentityService.Logger.IdentityEcsTextFormatter,
IdentityService
dis_serilog_path_format = {Путь до каталога для логов Identity Service внутри
контейнера}
dis_connection_strings_default_connection = User
ID=${pg_user};Password=${pg_password};Host=${pg_host};Port=${pg_port};Database=${p
g_database_dis};Pooling={Использовать пул подключений к БД};
dis_elastic_apm_server_urls = ${elastic_apm_server_urls}
dis_elastic_apm_service_name = ${dis_log_properties_service_name}
dis_elastic_apm_log_level = ${elastic_apm_log_level}
dis_elastic_apm_transaction_sample_rate = ${elastic_apm_transaction_sample_rate}
dis_elastic_apm_span_frames_min_duration = ${elastic_apm_span_frames_min_duration}
dis_elastic_apm_metrics_interval = ${elastic_apm_metric_interval}
dis_elastic_apm_capture_body = ${elastic_apm_capture_body}
dis_certificate_name = {Наименование файла сертификата}
dis_certificate_password = {Пароль сертификата}

```

[PublishingService](#)

```

dps_replicas = {Количество запускаемых копий сервиса Publishing Service}
dps_external_port = {TCP-порт сервиса Publishing Service}
dps_serilog_default = ${log_level}
dps_serilog_entity_framework_core = Error
dps_serilog_system = Error
dps_serilog_elastic_apm = Error
dps_serilog_application = PublishingService
dps_serilog_name = RollingFile
dps_serilog_formatter = PublishingService.Logger.LoggerEcsTextFormatter,
PublishingService
dps_serilog_path_format = {Путь до каталога для логов Publishing Service внутри
контейнера}
dps_elastic_apm_server_urls = ${elastic_apm_server_urls}
dps_elastic_apm_transaction_sample_rate = ${elastic_apm_transaction_sample_rate}
dps_elastic_apm_span_frames_min_duration = ${elastic_apm_span_frames_min_duration}
dps_elastic_apm_metrics_interval = ${elastic_apm_metric_interval}
dps_elastic_apm_capture_body = ${elastic_apm_capture_body}
dps_elastic_apm_capture_body_content_types = "*"
dps_queue_connection = ${rabbit_url}
dps_queues_message_ttl = ${queues_message_ttl}
dps_queues_services = {Список сервисов}
dps_queue_retry_connection_count = {Количество повторных попыток подключения к
RabbitMQ}

```

[TextExtractor](#)

```

dtes_replicas = {Количество запускаемых копий сервиса Text Extractor Service}
dtes_external_port = {TCP-порт сервиса Text Extractor Service}
dtes_serilog_default = ${log_level}
dtes_serilog_microsoft = Warning
dtes_serilog_aspnet_core = Information
dtes_serilog_entity_framework_core = Error
dtes_serilog_system = Error

```

```

dtes_serilog_elastic_apm = Error
dtes_serilog_application = TextExtractorService
dtes_serilog_name = RollingFile
dtes_serilog_formatter = TextExtractorService.Logger.SmartEcsTextFormatter,
TextExtractorService
dtes_serilog_path_format = {Путь до каталога для логов Text Extractor Service
внутри контейнера}
dtes_background_remover = {Алгоритм удаления фона}
dtes_extraction_quality = {Качество извлечения текстового слоя}
dtes_ocr_text_from_pdf = {Всегда извлекать текстовый слой из PDF-документов, даже
если текстовый слой уже существует}
dtes_ocr_dpi = {Разрешение, используемое для извлечения текста из PDF-документов
без текстового слоя или из документов, у которых игнорируется существующий
текстовый слой}
dtes_convert_fact_positions = {Учитывать поворот документов при получении
координат
фактов}
dtes_page_classifier_words_count = {Количество слов на странице, которые
используются для определения начала нового документа}
dtes_use_fixed_forms = {При обработке изображений применять жесткие формы
документов}
dtes_detect_orientation = {Определять ориентацию страниц}
dtes_find_objects = {Искать подписи}
dtes_find_stamps = {Искать печати}
dtes_prefetch_count = {Максимальное количество одновременно обрабатываемых
запросов}
dtes_parallel_processing_page_count = {Количество одновременно обрабатываемых
страниц документа}
dtes_max_pages = {Максимальное число страниц, при котором из документа будет
извлекаться текстовый слой}
dtes_debug_images_enabled = {Режим отладки для сохранения промежуточных
результатов обработки изображений}
dtes_debug_images_path = {Каталог для сохранения промежуточных результатов
обработки изображений внутри контейнера}
dtes_debug_images_path_host = {Каталог для сохранения промежуточных результатов
обработки изображений}
dtes_elastic_apm_server_urls = ${elastic_apm_server_urls}
dtes_elastic_apm_transaction_sample_rate = ${elastic_apm_transaction_sample_rate}
dtes_elastic_apm_span_frames_min_duration =
${elastic_apm_span_frames_min_duration}
dtes_elastic_apm_metrics_interval = ${elastic_apm_metric_interval}
dtes_elastic_apm_capture_body = ${elastic_apm_capture_body}
dtes_elastic_apm_capture_body_content_types = "*"
dtes_queues_message_ttl = ${queues_message_ttl}
dtes_queue_connection = ${rabbit_url}
dtes_queue_retry_connection_count = {Количество повторных попыток подключения к
RabbitMQ}

# DocumentDisassemblerService
ddds_replicas = {Количество запускаемых копий сервиса Directum Document
Disassembler Service}
ddds_external_port = {TCP-порт сервиса Directum Document Disassembler Service}
ddds_serilog_default = ${log_level}
ddds_serilog_microsoft = Warning
ddds_serilog_aspnet_core = Information
ddds_serilog_entity_framework_core = Error

```

```

ddds_serilog_system = Error
ddds_serilog_elastic_apm = Error
ddds_serilog_application = DocumentDisassemblerService
ddds_serilog_name = RollingFile
ddds_serilog_formatter = Logger.SmartEcsTextFormatter, Logger
ddds_serilog_path_format = {Путь до каталога для логов Directum Document
Disassembler Service внутри контейнера}
ddds_ocr_text_from_pdf = {Всегда извлекать текстовый слой из PDF-документов}
ddds_prefetch_count = {Максимальное количество одновременно обрабатываемых
запросов}
ddds_elastic_apm_server_urls = ${elastic_apm_server_urls}
ddds_elastic_apm_transaction_sample_rate = ${elastic_apm_transaction_sample_rate}
ddds_elastic_apm_span_frames_min_duration =
${elastic_apm_span_frames_min_duration}
ddds_elastic_apm_metrics_interval = ${elastic_apm_metric_interval}
ddds_elastic_apm_capture_body = ${elastic_apm_capture_body}
ddds_elastic_apm_capture_body_content_types = "*"
ddds_queues_message_ttl = ${queues_message_ttl}
ddds_queue_connection = ${rabbit_url}
ddds_queue_retry_connection_count = "{Количество повторных попыток подключения к
RabbitMQ}"

```

[DocumentAssemblerService](#)

```

ddas_replicas = {Количество запускаемых копий сервиса Directum Document Assembler
Service}
ddas_external_port = {TCP-порт сервиса Directum Document Assembler Service}
ddas_serilog_default = ${log_level}
ddas_serilog_microsoft = Warning
ddas_serilog_aspnet_core = Information
ddas_serilog_entity_framework_core = Error
ddas_serilog_system = Error
ddas_serilog_elastic_apm = Error
ddas_serilog_application = DocumentAssemblerService
ddas_serilog_name = RollingFile
ddas_serilog_formatter = Logger.SmartEcsTextFormatter, Logger
ddas_serilog_path_format = {Путь до каталога для логов Directum Document Assembler
Service внутри контейнера}
ddas_ocr_text_from_pdf = {Всегда извлекать текстовый слой из PDF-документов}
ddas_prefetch_count = {Максимальное количество одновременно обрабатываемых
запросов}
ddas_elastic_apm_server_urls = ${elastic_apm_server_urls}
ddas_elastic_apm_transaction_sample_rate = ${elastic_apm_transaction_sample_rate}
ddas_elastic_apm_span_frames_min_duration =
${elastic_apm_span_frames_min_duration}
ddas_elastic_apm_metrics_interval = ${elastic_apm_metric_interval}
ddas_elastic_apm_capture_body = ${elastic_apm_capture_body}
ddas_elastic_apm_capture_body_content_types = "*"
ddas_queues_message_ttl = ${queues_message_ttl}
ddas_queue_connection = ${rabbit_url}
ddas_queue_retry_connection_count = "{Количество повторных попыток подключения к
RabbitMQ}"

```

[FactExtractorService](#)

```

dfes_replicas = {Количество запускаемых копий сервиса Fact Extractor Service}

```

```
dfes_external_port = {TCP-порт сервиса Fact Extractor Service}
dfes_debug = {Записывать в лог-файл отладочные сообщения}
dfes_area = {Используемый раздел конфигурационного файла}
dfes_queue_connection = ${rabbit_url}
dfes_message_ttl_seconds = ${queues_message_ttl}
dfes_prefetch_count = {Максимальное количество одновременно обрабатываемых
запросов}
dfes_server_url = ${elastic_apm_server_urls}
dfes_metrics_interval = ${elastic_apm_metric_interval}
dfes_span_frames_min_duration = ${elastic_apm_span_frames_min_duration}
dfes_suffix = {Формат названия лог-файлов}
```

[FactExtractorRulesService](#)

```
dfesr_replicas = {Количество запускаемых копий сервиса Fact Extractor Rules
Service}
dfesr_external_port = {TCP-порт сервиса Fact Extractor Rules Service}
dfesr_debug = {Записывать в лог-файл отладочные сообщения}
dfesr_area = {Используемый раздел конфигурационного файла}
dfesr_parser_file_name = {Имя файла синтаксического анализатора для извлечения
фактов}
dfesr_parser_debug = {Записывать в лог-файл отладочные сообщения в процессе
извлечения фактов и расчета вероятностей}
dfesr_parsers_count = {Количество экземпляров синтаксического анализатора в пуле}
dfesr_parser_timeout = {Таймаут в секундах, после которого обработка запроса
прерывается}
dfesr_queue_connection = ${rabbit_url}://?heartbeat={Значение}
dfesr_message_ttl_seconds = ${queues_message_ttl}
dfesr_prefetch_count = {Максимальное количество одновременно обрабатываемых
запросов}
dfesr_server_url = ${elastic_apm_server_urls}
dfesr_metrics_interval = ${elastic_apm_metric_interval}
dfesr_span_frames_min_duration = ${elastic_apm_span_frames_min_duration}
dfesr_suffix = {Формат названия лог-файлов}
```

[FactExtractorLearnService](#)

```
dfesl_replicas = {Количество запускаемых копий сервиса Fact Extractor Learn
Service}
dfesl_max_replicas_per_node = {Ограничение количества запускаемых копий сервиса
Fact Extractor Learn Service на одном сервере}
dfesl_external_port = {TCP-порт сервиса Fact Extractor Learn Service}
dfesl_debug = {Записывать в лог-файл отладочные сообщения}
dfesl_worker_cache_lifetime = {Время хранения результата обучения в кэше сервиса в
секундах}
dfesl_can_train_models = {Разрешить использовать сервис как сервис обучения}
dfesl_dbname = ${pg_database}
dfesl_dbuser = ${pg_user}
dfesl_dbhost = ${pg_host}
dfesl_dbpass = ${pg_password}
dfesl_dbport = ${pg_port}
dfesl_queue_connection = ${rabbit_url}://?heartbeat={Значение}
dfesl_message_ttl_seconds = ${queues_message_ttl}
dfesl_prefetch_count = {Максимальное количество одновременно обрабатываемых
запросов}
dfesl_model_cache_lifetime = {Время хранения модели в пуле}
dfesl_subprocesses = {Количество подпроцессов}
dfesl_algorithm = {Алгоритм обучения}
```

```

dfesl_c1 = {Коэффициент L1-регуляризации}
dfesl_c2 = {Коэффициент L2-регуляризации}
dfesl_epsilon = {Точность вычисления показателя качества модели}
dfesl_delta = {Относительный показатель качества модели}
dfesl_slice_period = {Количество итераций обучения, через которое оценивается
относительный показатель качества модели}
dfesl_max_iterations = {Максимальное число итераций обучения}
dfesl_all_possible_transitions = {Генерировать все сочетания классов}
dfesl_logging_period = {Интервал логирования процесса обучения в итерациях}
dfesl_test_size = {Размер выборки для расчета метрик}
dfesl_server_url = ${elastic_apm_server_urls}
dfesl_metrics_interval = ${elastic_apm_metric_interval}
dfesl_span_frames_min_duration = ${elastic_apm_span_frames_min_duration}
dfesl_suffix = {Формат названия лог-файлов}

```

[ObjectDetectionService](#)

```

dods_replicas = {Количество запускаемых копий сервиса Object Detection Service}
dods_external_port = {TCP-порт сервиса Fact Extractor Learn Service}
dods_debug = {Записывать в лог-файл отладочные сообщения}
dods_queue_connection = ${rabbit_url}
dods_message_ttl_seconds = ${queues_message_ttl}
dods_prefetch_count = {Максимальное количество одновременно обрабатываемых
запросов}
dods_server_url = ${elastic_apm_server_urls}
dods_metrics_interval = ${elastic_apm_metric_interval}
dods_span_frames_min_duration = ${elastic_apm_span_frames_min_duration}
dods_suffix = {Формат названия лог-файлов}

```

[ImageClassifierService](#)

```

dics_replicas = {Количество запускаемых копий сервиса Image Classifier Service}
dics_external_port = {TCP-порт сервиса Image Classifier Service}
dics_debug = {Записывать в лог-файл отладочные сообщения}
dics_area = {Используемый раздел конфигурационного файла}
dics_queue_connection = ${rabbit_url}
dics_message_ttl_seconds = ${queues_message_ttl}
dics_prefetch_count = {Максимальное количество одновременно обрабатываемых
запросов}
dics_multithreaded_predict = {Использовать многопоточность}
dics_server_url = ${elastic_apm_server_urls}
dics_metrics_interval = ${elastic_apm_metric_interval}
dics_span_frames_min_duration = ${elastic_apm_span_frames_min_duration}
dics_suffix = {Формат названия лог-файлов}

```

[TextClassifierService](#)

```

dtcs_replicas = {Количество запускаемых копий сервиса Text Classifier Service}
dtcs_external_port = {TCP-порт сервиса Text Classifier Service}
dtcs_debug = {Записывать в лог-файл отладочные сообщения}
dtcs_area = {Используемый раздел конфигурационного файла}
dtcs_max_models_in_pool = {Максимальное количество моделей классификации в пуле}
dtcs_model_cache_lifetime = {Время хранения модели классификации в пуле}
dtcs_dbname = ${pg_database}
dtcs_dbuser = ${pg_user}
dtcs_dbhost = ${pg_host}
dtcs_dbpass = ${pg_password}
dtcs_dbport = ${pg_port}

```

```
dtcs_csv_text_field = {Название поля с текстом для обучения данными в формате CSV}
dtcs_csv_class_field = {Название поля с классами для обучения данными в формате CSV многоклассового классификатора}
dtcs_csv_classes_field = {Название поля с классами для обучения данными в формате CSV многозначного классификатора}
dtcs_csv_external_id_field = {Название поля с внешним идентификатором класса для обучения данными в формате CSV многоклассового классификатора}
dtcs_repository = {Тип репозитория для хранения моделей классификации}
dtcs_can_train_models = {Разрешить использовать сервис как сервис обучения классификаторов}
dtcs_queue_connection = ${rabbit_url}
dtcs_message_ttl_seconds = ${queues_message_ttl}
dtcs_prefetch_count = {Максимальное количество одновременно обрабатываемых запросов}
dtcs_max_text_extraction_requests = {Количество документов для одновременного извлечения текста при обучении классификатора}
dtcs_allow_local_path = {Разрешить обучение из локальной папки}
dtcs_server_url = ${elastic_apm_server_urls}
dtcs_metrics_interval = ${elastic_apm_metric_interval}
dtcs_span_frames_min_duration = ${elastic_apm_span_frames_min_duration}
dtcs_suffix = {Формат названия лог-файлов}
```

Smart Service

Сервис Smart Service настраивается в одноименной секции конфигурационного файла `.env`.

Структура секции:

```
# SmartService
dss_replicas = {Количество запускаемых копий сервиса Smart Service}
dss_address = {Адрес сервиса Smart Service}
dss_external_port = {TCP-порт сервиса Smart Service}
dss_connection_strings_default_connection = ${pg_connection}
dss_serilog_default = ${log_level}
dss_serilog_entity_framework_core = Error
dss_serilog_system = Error
dss_serilog_elastic_apm = Error
dss_serilog_application = SmartService
dss_serilog_name = RollingFile
dss_serilog_formatter = SmartService.Logger.SmartEcsTextFormatter, SmartService
dss_serilog_path_format = {Путь до каталога для логов Smart Service внутри
контейнера}
dss_identity_service = {Адрес сервиса Identity Service}
dss_allow_local_path = {Разрешить обучение классификаторов из локального каталога}
dss_allow_anonymous_access = {Разрешить анонимную работу}
dss_anonymous_client_id = {Логин анонимного клиента}
dss_anonymous_client_secret = {Пароль анонимного клиента}
dss_storage_endpoint = {Адрес сервиса хранения объектов}
dss_storage_access_key = ${storage_service_access_key}
dss_storage_secret_key = ${storage_service_secret_key}
dss_storage_bucket = {Имя контейнера}
dss_storage_uri_lifetime_seconds = {Время хранения ссылок на объекты в секундах}
dss_storage_file_lifetime_hours = {Время хранения объектов в часах}
dss_storage_retry_connection_count = {Количество повторных попыток подключения к
Storage Service}
dss_elastic_apm_server_urls = ${elastic_apm_server_urls}
dss_elastic_apm_transaction_sample_rate = ${elastic_apm_transaction_sample_rate}
dss_elastic_apm_span_frames_min_duration = ${elastic_apm_span_frames_min_duration}
dss_elastic_apm_metrics_interval = ${elastic_apm_metric_interval}
dss_elastic_apm_capture_body = ${elastic_apm_capture_body}
dss_elastic_apm_capture_body_content_types = "*"
dss_queues_message_ttl = ${queues_message_ttl}
dss_queue_connection = ${rabbit_url}
dss_queue_retry_connection_count = {Количество повторных попыток подключения к
RabbitMQ}
```

- **dss_replicas.** Количество запускаемых копий сервиса Smart Service. Возможные значения: **0** – сервис DSS не будет запущен при следующем запуске сервисов, **1** – сервис DSS будет запущен при следующем запуске сервисов. Значение по умолчанию **1**;
- **dss_address.** Адрес контейнера с сервисом Smart Service. Значение по умолчанию **http://dss:80**;
- **dss_external_port.** TCP-порт сервиса Smart Service для обращения к контейнеру через внешнее подключение к хосту. Значение по умолчанию **61100**;

- **dss_connection_strings_default_connection.** Строка подключения к базе данных PostgreSQL. Подставляется значение, указанное для параметра **pg_connection** в конфигурационном файле .env в [секции PostgreSQL](#);
- **dss_serilog_default.** Параметры ведения лог-файлов;
- **dss_serilog_path_format.** Путь до каталога с лог-файлами Smart Service;
- **dss_identity_service.** Адрес сервиса аутентификации и авторизации пользователей Identity Service. Пример значения **http://localhost:9030**;
- **dss_allow_local_path.** Разрешить обучение классификаторов из локального каталога сервера. Возможные значения: **True, False**. Если установлено значение **False**, повышается безопасность работы. Значение по умолчанию **True**;
- **dss_allow_anonymous_access.** Разрешить анонимную работу системы. Возможные значения: **True, False**. Значение **True** устанавливается при локальной поставке сервисов Ario и позволяет не использовать аутентификацию. При облачной поставке устанавливается значение **False**, после чего требуется обязательная аутентификация и передача токена в каждом запросе;
- **dss_anonymous_client_id.** Логин анонимного клиента, который взаимодействует с сервисами Ario при разрешенной анонимной работе. Значение по умолчанию **anonymous**;
- **dss_anonymous_client_secret.** Пароль анонимного клиента, который взаимодействует с сервисами Ario при разрешенной анонимной работе. Значение по умолчанию **ArioAnonymous**;
- **dss_storage_endpoint.** Веб-адрес сервиса хранения объектов. Указывается без протокола http/https. Пример значения: **localhost:9000**;
- **dss_storage_access_key.** Имя пользователя для подключения к сервису хранения объектов. Подставляется значение, указанное для данного параметра в конфигурационном файле .env в [секции Storage Service](#);
- **dss_storage_secret_key.** Пароль для подключения к сервису хранения объектов. Подставляется значение, указанное для данного параметра в конфигурационном файле .env в [секции Storage Service](#);
- **dss_storage_bucket.** Имя каталога для хранения объектов. Значение по умолчанию **ario**;
- **dss_storage_uri_lifetime_seconds.** Время хранения ссылок на объекты в секундах. Значение по умолчанию **3600**;
- **dss_storage_file_lifetime_hours.** Время хранения объектов в хранилище в часах. Значение по умолчанию **72**;
- **dss_storage_retry_connection_count.** Количество повторных попыток подключения к сервису хранения объектов. Значение по умолчанию **5**;
- **dss_elastic_apm_server_urls.** Адрес сервиса Elastic APM Service. Подставляется значение, указанное для данного параметра в конфигурационном файле .env в [секции Elastic APM Service](#);
- **dss_elastic_apm_transaction_sample_rate.** Регулярность отправки трейсов в Elastic APM Service. Подставляется значение, указанное для данного параметра в конфигурационном файле .env в секции Elastic APM Service;

- **dss_elastic_apm_span_frames_min_duration.** Если при обработке документа возникла ошибка и время обработки превышает значение параметра, в лог-файл записывается не только факт ошибки, но и подробная информация о ней. Используется значение, указанное для этого параметра в конфигурационном файле .env в [секции Elastic APM Service](#);
- **dss_elastic_apm_metrics_interval.** Интервал сбора метрик о состоянии памяти и CPU. Подставляется значение, указанное для данного параметра в конфигурационном файле .env в [секции Elastic APM Service](#);
- **dss_elastic_apm_capture_body.** Сообщения, которые передаются в сервис Elastic APM Service. Используется значение, указанное для этого параметра в конфигурационном файле .env в [секции Elastic APM Service](#);
- **dss_queue_message_ttl.** Время хранения сообщений об обработке документа в RabbitMQ в секундах. Подставляется значение, указанное для [данного параметра](#) в конфигурационном файле .env;
- **dss_queue_connection.** Строка подключения к очереди сообщений RabbitMQ. Подставляется значение, указанное для параметра **rabbit_url** в конфигурационном файле .env в [секции RabbitMQ](#);
- **dss_queue_retry_connection_count.** Количество повторных попыток подключения DSS к брокеру сообщений. Значение по умолчанию **5**.

Identity Service

Сервис Identity Service настраивается в одноименной секции конфигурационного файла .env.

Структура секции:

```
# IdentityService
dis_replicas = {Количество запускаемых копий сервиса Identity Service}
dis_external_port = {TCP-порт сервиса Identity Service}
dis_serilog_default = ${log_level}
dis_serilog_entity_framework_core = Error
dis_serilog_system = Error
dis_serilog_elastic_apm = Error
dis_serilog_application = IdentityService
dis_serilog_name = RollingFile
dis_serilog_formatter = IdentityService.Logger.IdentityEcsTextFormatter,
IdentityService
dis_serilog_path_format = {Путь до каталога для логов Identity Service внутри
контейнера}
dis_connection_strings_default_connection = User
ID=${pg_user};Password=${pg_password};Host=${pg_host};Port=${pg_port};Database=${p
g_database_dis};Pooling={Использовать пул подключений к БД};
dis_elastic_apm_server_urls = ${elastic_apm_server_urls}
dis_elastic_apm_service_name = ${dis_log_properties_service_name}
dis_elastic_apm_log_level = ${elastic_apm_log_level}
dis_elastic_apm_transaction_sample_rate = ${elastic_apm_transaction_sample_rate}
dis_elastic_apm_span_frames_min_duration = ${elastic_apm_span_frames_min_duration}
dis_elastic_apm_metrics_interval = ${elastic_apm_metric_interval}
dis_elastic_apm_capture_body = ${elastic_apm_capture_body}
dis_certificate_name = {Наименование файла сертификата}
dis_certificate_password = {Пароль сертификата}
```

- **dis_replicas**. Количество запускаемых копий сервиса Identity Service. Возможные значения: **0** – сервис DIS не будет запущен при следующем запуске сервисов, **1** – сервис DIS будет запущен при следующем запуске сервисов. Значение по умолчанию **1**;
- **dis_external_port**. TCP-порт сервиса Identity Service для обращения к контейнеру через внешнее подключение к хосту. Значение по умолчанию **9030**;
- **dis_serilog_default**. Параметры ведения лог-файлов;
- **dis_serilog_path_format**. Путь до каталога с лог-файлами Identity Service;
- **dis_connection_strings_default_connection** – параметры подключения к базе данных PostgreSQL, где:
 - **UserID** и **Password** – имя и пароль пользователя для подключения к СУБД PostgreSQL. Хранятся в зашифрованном виде;
 - **Host** – адрес СУБД PostgreSQL;
 - **Port** – порт для подключения к СУБД PostgreSQL;
 - **Database** – имя базы данных сервиса Identity Service;
 - **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**.

ПРИМЕЧАНИЕ. Identity Service имеет собственную базу данных, отличную от базы данных Smart Service;

- **dis_elastic_apm_server_urls**. Адрес сервиса Elastic APM Service. Подставляется значение, указанное для данного параметра в конфигурационном файле .env в [секции Elastic APM Service](#);
- **dis_elastic_apm_service_name**. Имя, под которым сервис Identity Service отображается в лог-файлах Elastic APM Service. Значение по умолчанию **Identity Service**;
- **dis_elastic_apm_log_level**. Уровень логирования библиотеки, которая отправляет в сервис Elastic APM Service информацию о запросах и метриках сервиса Identity Service. Значение по умолчанию **Error**;
- **dis_elastic_apm_transaction_sample_rate**. Регулярность отправки трейсов в Elastic APM Service. Подставляется значение, указанное для данного параметра в конфигурационном файле .env в [секции Elastic APM Service](#);
- **dis_elastic_apm_span_frames_min_duration**. Если при обработке документа возникла ошибка и время обработки превышает значение параметра, в лог-файл записывается не только факт ошибки, но и подробная информация о ней. Используется значение, указанное для этого параметра в конфигурационном файле .env в [секции Elastic APM Service](#);
- **dis_elastic_apm_metrics_interval**. Интервал сбора метрик о состоянии памяти и CPU. Подставляется значение, указанное для данного параметра в конфигурационном файле .env в [секции Elastic APM Service](#);
- **dis_elastic_apm_capture_body**. Сообщения, которые передаются в сервис Elastic APM Service. Используется значение, указанное для этого параметра в конфигурационном файле .env в [секции Elastic APM Service](#);

- **dis_certificate_name.** Наименование файла сертификата для выдачи токена пользователю;
- **dis_certificate_password.** Пароль пользователя сертификата.

Publishing Service

Сервис Publishing Service настраивается в одноименной секции конфигурационного файла .env.

Структура секции:

```
# PublishingService
dps_replicas = {Количество запускаемых копий сервиса Publishing Service}
dps_external_port = {TCP-порт сервиса Publishing Service}
dps_serilog_default = ${log_level}
dps_serilog_entity_framework_core = Error
dps_serilog_system = Error
dps_serilog_elastic_apm = Error
dps_serilog_application = PublishingService
dps_serilog_name = RollingFile
dps_serilog_formatter = PublishingService.Logger.LoggerEcsTextFormatter,
PublishingService
dps_serilog_path_format = {Путь до каталога для логов Publishing Service внутри
контейнера}
dps_elastic_apm_server_urls = ${elastic_apm_server_urls}
dps_elastic_apm_transaction_sample_rate = ${elastic_apm_transaction_sample_rate}
dps_elastic_apm_span_frames_min_duration = ${elastic_apm_span_frames_min_duration}
dps_elastic_apm_metrics_interval = ${elastic_apm_metric_interval}
dps_elastic_apm_capture_body = ${elastic_apm_capture_body}
dps_elastic_apm_capture_body_content_types = "*"
dps_queue_connection = ${rabbit_url}
dps_queues_message_ttl = ${queues_message_ttl}
dps_queues_services = {Список сервисов}
dps_queue_retry_connection_count = {Количество повторных попыток подключения к
RabbitMQ}
```

- **dps_replicas.** Количество запускаемых копий сервиса Publishing Service. Возможные значения: **0** – сервис DPS не будет запущен при следующем запуске сервисов, **1** – сервис DPS будет запущен при следующем запуске сервисов. Значение по умолчанию **1**;
- **dps_external_port.** TCP-порт сервиса Publishing Service для обращения к контейнеру через внешнее подключение к хосту. Значение по умолчанию **9057**;
- **dps_serilog_default.** Параметры ведения лог-файлов;
- **dps_serilog_path_format.** Путь до каталога с лог-файлами Publishing Service;
- **dps_elastic_apm_server_urls.** Адрес сервиса Elastic APM Service. Подставляется значение, указанное для данного параметра в конфигурационном файле .env в [секции Elastic APM Service](#);
- **dps_elastic_apm_transaction_sample_rate.** Регулярность отправки трейсов в Elastic APM Service. Подставляется значение, указанное для данного параметра в конфигурационном файле .env в [секции Elastic APM Service](#);

- **dps_elastic_apm_span_frames_min_duration.** Если при обработке документа возникла ошибка и время обработки превышает значение параметра, в лог-файл записывается не только факт ошибки, но и подробная информация о ней. Используется значение, указанное для этого параметра в конфигурационном файле .env в секции Elastic APM Service;
- **dps_elastic_apm_metrics_interval.** Интервал сбора метрик о состоянии памяти и CPU. Подставляется значение, указанное для данного параметра в конфигурационном файле .env в [секции Elastic APM Service](#);
- **dps_elastic_apm_capture_body.** Сообщения, которые передаются в сервис Elastic APM Service. Используется значение, указанное для этого параметра в конфигурационном файле .env в [секции Elastic APM Service](#);
- **dps_queue_connection.** Строка подключения к очереди сообщений RabbitMQ. Подставляется значение, указанное для параметра **rabbit_url** в конфигурационном файле .env в [секции RabbitMQ](#);
- **dps_queue_message_ttl.** Время хранения сообщений об обработке документа в RabbitMQ в секундах. Подставляется значение, указанное для [данного параметра](#) в конфигурационном файле .env;
- **dps_queue_services.** Список сервисов, с которыми работает сервис Publishing Service. Если сервис не указан, то он и его дочерние сервисы не обновляются. Значение по умолчанию **Fact Extractor Service, Fact Extractor Rules Service, Fact Extractor Learn Service**;
- **dps_queue_retry_connection_count.** Количество повторных попыток подключения DSS к брокеру сообщений. Значение по умолчанию **5**.

Text Extractor Service

Сервис Text Extractor Service настраивается в одноименной секции конфигурационного файла .env.

Структура секции:

```
# TextExtractor
dtes_replicas = {Количество запускаемых копий сервиса Text Extractor Service}
dtes_external_port = {TCP-порт сервиса Text Extractor Service}
dtes_serilog_default = ${log_level}
dtes_serilog_microsoft = Warning
dtes_serilog_aspnet_core = Information
dtes_serilog_entity_framework_core = Error
dtes_serilog_system = Error
dtes_serilog_elastic_apm = Error
dtes_serilog_application = TextExtractorService
dtes_serilog_name = RollingFile
dtes_serilog_formatter = TextExtractorService.Logger.SmartEcsTextFormatter,
TextExtractorService
dtes_serilog_path_format = {Путь до каталога для логов Text Extractor Service
внутри контейнера}
dtes_background_remover = {Алгоритм удаления фона}
dtes_extraction_quality = {Качество извлечения текстового слоя}
dtes_ocr_text_from_pdf = {Всегда извлекать текстовый слой из PDF-документов}
dtes_ocr_dpi = {Разрешение, используемое для извлечения текста из PDF-документов
без текстового слоя или из документов, у которых игнорируется существующий
```

```

текстовый слой}
dtes_convert_fact_positions = {Учитывать поворот документов при получении
координат
фактов}
dtes_page_classifier_words_count = {Количество слов на странице, которые
используются для определения начала нового документа}
dtes_use_fixed_forms = {Использовать жесткие формы документов}
dtes_detect_orientation = {Определять ориентацию страниц}
dtes_find_objects = {Искать подписи}
dtes_find_stamps = {Искать печати}
dtes_prefetch_count = {Максимальное количество одновременно обрабатываемых
запросов}
dtes_parallel_processing_page_count = {Количество одновременно обрабатываемых
страниц документа}
dtes_max_pages = {Максимальное число страниц, при котором из документа будет
извлекаться текстовый слой}
dtes_debug_images_enabled = {Режим отладки для сохранения промежуточных
результатов обработки изображений}
dtes_debug_images_path = {Каталог для сохранения промежуточных результатов
обработки изображений внутри контейнера}
dtes_debug_images_path_host = {Каталог для сохранения промежуточных результатов
обработки изображений}
dtes_elastic_apm_server_urls = ${elastic_apm_server_urls}
dtes_elastic_apm_transaction_sample_rate = ${elastic_apm_transaction_sample_rate}
dtes_elastic_apm_span_frames_min_duration =
${elastic_apm_span_frames_min_duration}
dtes_elastic_apm_metrics_interval = ${elastic_apm_metric_interval}
dtes_elastic_apm_capture_body = ${elastic_apm_capture_body}
dtes_elastic_apm_capture_body_content_types = "*"
dtes_queues_message_ttl = ${queues_message_ttl}
dtes_queue_connection = ${rabbit_url}
dtes_queue_retry_connection_count = {Количество повторных попыток подключения к
RabbitMQ}

```

- **dtes_replicas.** Количество запускаемых копий сервиса Text Extractor Service. Возможные значения: **0** – сервис DTES не будет запущен при следующем запуске сервисов, **1** – сервис DTES будет запущен при следующем запуске сервисов. Значение по умолчанию **1**;
- **dtes_external_port.** TCP-порт сервиса Text Extractor Service для обращения к контейнеру через внешнее подключение к хосту. Значение по умолчанию **9027**;
- **dtes_serilog_default.** Параметры ведения лог-файлов;
- **dtes_serilog_path_format.** Путь до каталога с лог-файлами Text Extractor Service;
- **dtes_background_remover.** Алгоритм, который используется для удаления фона в документе. Возможные значения:
 - **ConnectedComponents.** Основан на методе поиска связанных компонент установленного размера и последующем удалении компонент меньше этого значения. Быстрее и точнее обрабатывает простые и сложные фоны, но иногда может задевать текстовый слой: стирать границы таблиц и удалять знаки препинания;

- **Sobel.** Основан на методе поиска границ изображения с помощью оператора Собеля. Обработывает простые фоны;
- **None.** Фон не удаляется.

Значение по умолчанию **ConnectedComponents**;

- **dtes_extraction_quality.** Качество извлечения текстового слоя. Возможные значения:
 - **fast.** Высокая скорость извлечения, при этом снижается качество;
 - **quality.** Высокое качество извлечения, при этом снижается скорость.

Значение по умолчанию **fast**;

- **dtes_ocr_text_from_pdf.** Всегда извлекать текстовый слой из PDF-документов, даже если текстовый слой уже существует. Возможные значения: **True**, **False**. Значение по умолчанию **True**;
- **dtes_ocr_dpi.** 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**;
- **dtes_convert_fact_positions.** 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**;
- **dtes_page_classifier_words_count.** Количество слов на странице, которые используются для определения начала нового документа. Укажите целое число больше нуля. Значение по умолчанию **50**;
- **dtes_use_fixed_forms.** 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**;
- **dtes_detect_orientation.** Определять ориентацию страниц. Возможные значения: **True**, **False**. Значение по умолчанию **True**. Если нет необходимости в определении ориентации страниц, для ускорения обработки документов установите значение **False**;
- **dtes_find_objects.** Искать подписи. Возможные значения: **True**, **False**. Значение по умолчанию **True**. Если нет необходимости в поиске, для ускорения обработки документов установите значение **False**;
- **dtes_find_stamps.** Искать печати. Возможные значения: **True**, **False**. Значение по умолчанию **True**. Если нет необходимости в поиске, для ускорения обработки документов установите значение **False**;
- **dtes_prefetch_count.** Максимальное количество одновременно обрабатываемых сервисом запросов. Значение по умолчанию **0**. Если установлено такое значение, то оптимальное количество одновременно обрабатываемых документов подбирается автоматически. Оно зависит от доступного ресурса системы.

Доступный ресурс представляет собой число логических ядер процессора (за вычетом одного ядра на работу ОС). Формула для расчета: **CPU-1**;

- **dtes_parallel_processing_page_count.** Количество одновременно обрабатываемых страниц документа. Значение по умолчанию **0**. Если установлено такое значение, то оптимальное количество одновременно обрабатываемых страниц подбирается автоматически. Оно зависит от доступного ресурса системы, который рассчитывается по формуле, аналогичной расчету ресурса для параметра **dtes_prefetch_count**. При этом, если значение доступного ресурса:
 - больше 1, то количество одновременно обрабатываемых страниц равно 2;
 - меньше или равно 1, то количество одновременно обрабатываемых страниц равно 1;
- **dtes_max_pages.** Максимальное число страниц, при котором из документа извлекается текстовый слой. Если количество страниц превышает установленный лимит, то текстовый слой не извлекается.

Если из документа уже был извлечен текстовый слой, то при конвертации в формат PDF он сохранится. Исключение составляют многостраничные файлы в формате TIFF, при конвертации которых в PDF текстовый слой не сохраняется. Также такие файлы не разделяются на несколько документов.

Значение по умолчанию **100**. Если указать в параметре **0** или отрицательное число, то лимит будет отключен;

- **dtes_debug_images_enabled.** Режим отладки для сохранения промежуточных результатов обработки изображений. Возможные значения: **True**, **False**. Если указано значение **True**, то во временный каталог, указанный в параметре **dtes_debug_images_path**, сохраняются промежуточные результаты обработки изображений на каждом из этапов. Значение по умолчанию **False**.

Чтобы включить режим отладки:

1. В параметре **dtes_debug_images_enabled** установите значение **True**.

2. В конфигурационном файле `deploy.yml` в секции **dtes** в разделе **volumes** раскомментируйте строку:

```
#- ${dtes_debug_images_path_host}:/usr/src/textextractor-  
service/${dtes_debug_images_path}
```

3. Перезапустите сервисы Ario. Для этого последовательно выполните команды:

```
./do.sh ario down  
./do.sh ario up
```

- **dtes_debug_images_path.** Адрес временного каталога для сохранения промежуточных результатов обработки изображений. Значение по умолчанию **tmp/debug**. Не рекомендуется менять значение параметра;
- **dtes_debug_images_path_host.** Каталог для сохранения промежуточных результатов обработки изображений на сервере. Значение по умолчанию **./etc/textextractor-service/debug**. Не рекомендуется менять значение параметра;
- **dtes_elastic_apm_server_urls.** Адрес сервиса Elastic APM Service. Подставляется значение, указанное для данного параметра в конфигурационном файле `.env` в [секции Elastic APM Service](#);
- **dtes_elastic_apm_transaction_sample_rate.** Регулярность отправки трейсов в Elastic APM Service. Подставляется значение, указанное для данного параметра в конфигурационном файле `.env` в [секции Elastic APM Service](#);

- **dtes_elastic_apm_span_frames_min_duration.** Если при обработке документа возникла ошибка и время обработки превышает значение параметра, в лог-файл записывается не только факт ошибки, но и подробная информация о ней. Используется значение, указанное для этого параметра в конфигурационном файле `.env` в секции `Elastic APM Service`;
- **dtes_elastic_apm_metrics_interval.** Интервал сбора метрик о состоянии памяти и CPU. Подставляется значение, указанное для данного параметра в конфигурационном файле `.env` в [секции Elastic APM Service](#);
- **dtes_elastic_apm_capture_body.** Сообщения, которые передаются в сервис Elastic APM Service. Используется значение, указанное для этого параметра в конфигурационном файле `.env` в [секции Elastic APM Service](#);
- **dtes_queue_message_ttl.** Время хранения сообщений об обработке документа в RabbitMQ в секундах. Подставляется значение, указанное для [данного параметра](#) в конфигурационном файле `.env`;
- **dtes_queue_connection.** Строка подключения к очереди сообщений RabbitMQ. Подставляется значение, указанное для параметра `rabbit_url` в конфигурационном файле `.env` в [секции RabbitMQ](#);
- **dtes_retry_connection_count.** Количество повторных попыток подключения к брокеру сообщений. Значение по умолчанию `5`.

Document Disassembler Service

Сервис Document Disassembler Service настраивается в одноименной секции конфигурационного файла `.env`.

Структура секции:

```
# DocumentDisassemblerService
ddds_replicas = {Количество запускаемых копий сервиса Directum Document
Disassembler Service}
ddds_external_port = {TCP-порт сервиса Directum Document Disassembler Service}
ddds_serilog_default = ${log_level}
ddds_serilog_microsoft = Warning
ddds_serilog_aspnet_core = Information
ddds_serilog_entity_framework_core = Error
ddds_serilog_system = Error
ddds_serilog_elastic_apm = Error
ddds_serilog_application = DocumentDisassemblerService
ddds_serilog_name = RollingFile
ddds_serilog_formatter = Logger.SmartEcsTextFormatter, Logger
ddds_serilog_path_format = {Путь до каталога для логов Directum Document
Disassembler Service внутри контейнера}
ddds_ocr_text_from_pdf = {Всегда извлекать текстовый слой из PDF-документов}
ddds_prefetch_count = {Максимальное количество одновременно обрабатываемых
запросов}
ddds_elastic_apm_server_urls = ${elastic_apm_server_urls}
ddds_elastic_apm_transaction_sample_rate = ${elastic_apm_transaction_sample_rate}
ddds_elastic_apm_span_frames_min_duration = {elastic_apm_span_frames_min_duration}
ddds_elastic_apm_metrics_interval = ${elastic_apm_metric_interval}
ddds_elastic_apm_capture_body = ${elastic_apm_capture_body}
```

```

ddds_elastic_apm_capture_body_content_types = "*"
ddds_queues_message_ttl = ${queues_message_ttl}
ddds_queue_connection = ${rabbit_url}
ddds_queue_retry_connection_count = {Количество повторных попыток подключения к RabbitMQ}

```

- **ddds_replicas.** Количество запускаемых копий сервиса Directum Document Disassembler Service (DDDS). Возможные значения: **0** – сервис DDDS не будет запущен при следующем запуске сервисов, **1** – сервис DDDS будет запущен при следующем запуске сервисов. Значение по умолчанию **1**;
- **ddds_external_port.** TCP-порт сервиса Directum Document Disassembler Service для обращения к контейнеру через внешнее подключение к хосту. Значение по умолчанию **9057**;
- **ddds_serilog_default.** Параметры ведения лог-файлов.
Подробнее о значениях параметров см. на сервисе GitHub в [документации](#) библиотеки Serilog;
- **ddds_serilog_path_format.** Путь до каталога с лог-файлами Document Disassembler Service;
- **ddds_ocr_text_from_pdf.** Всегда извлекать текстовый слой из PDF-документов, даже если текстовый слой уже существует. Возможные значения: **True**, **False**. Значение по умолчанию **True**;
- **ddds_prefetch_count.** Максимальное количество одновременно обрабатываемых сервисом запросов. Значение по умолчанию **0**, при котором оптимальное количество одновременно обрабатываемых документов подбирается автоматически в зависимости от доступного ресурса системы.
ПРИМЕЧАНИЕ. Доступный ресурс представляет собой число логических ядер процессора (за вычетом одного ядра на работу ОС). Формула для расчета: **CPU-1**. Значение самого параметра вычисляется как половина от доступного ресурса с округлением вниз. Формула для расчета: **Floor((CPU-1)/2)**;
- **ddds_elastic_apm_server_urls.** Адрес сервиса Elastic APM Service. Используется значение, указанное для этого параметра в конфигурационном файле .env в [секции Elastic APM Service](#);
- **ddds_elastic_apm_transaction_sample_rate.** Регулярность отправки трейсов в Elastic APM Service. Используется значение, указанное для этого параметра в конфигурационном файле .env в [секции Elastic APM Service](#);
- **ddds_elastic_apm_span_frames_min_duration.** Если при обработке документа возникла ошибка и время обработки превышает значение параметра, в лог-файл записывается не только факт ошибки, но и подробная информация о ней. Используется значение, указанное для этого параметра в конфигурационном файле .env в [секции Elastic APM Service](#);
- **ddds_elastic_apm_metrics_interval.** Интервал сбора метрик о состоянии памяти и CPU. Используется значение, указанное для этого параметра в конфигурационном файле .env в [секции Elastic APM Service](#);
- **ddds_elastic_apm_capture_body.** Сообщения, которые передаются в сервис Elastic APM Service. Используется значение, указанное для этого параметра в конфигурационном файле .env в [секции Elastic APM Service](#);

- **ddds_queues_message_ttl.** Время хранения сообщений об обработке документа в RabbitMQ в секундах. Используется значение, указанное для [этого параметра](#) в конфигурационном файле .env;
- **ddds_queue_connection.** Строка подключения к очереди сообщений RabbitMQ. Используется значение, указанное для параметра **rabbit_url** в конфигурационном файле .env в [секции RabbitMQ](#);
- **ddds_queue_retry_connection_count.** Количество повторных попыток подключения DSS к брокеру сообщений. Значение по умолчанию **5**.

Document Assembler Service

Сервис Document Assembler Service настраивается в одноименной секции конфигурационного файла .env.

Структура секции:

```
# DocumentAssemblerService
ddas_replicas = {Количество запускаемых копий сервиса Directum Document Assembler Service}
ddas_external_port = {TCP-порт сервиса Directum Document Assembler Service}
ddas_serilog_default = ${log_level}
ddas_serilog_microsoft = Warning
ddas_serilog_aspnet_core = Information
ddas_serilog_entity_framework_core = Error
ddas_serilog_system = Error
ddas_serilog_elastic_apm = Error
ddas_serilog_application = DocumentAssemblerService
ddas_serilog_name = RollingFile
ddas_serilog_formatter = Logger.SmartEcsTextFormatter, Logger
ddas_serilog_path_format = {Путь до каталога для логов Directum Document Assembler Service внутри контейнера}
ddas_ocr_text_from_pdf = {Всегда извлекать текстовый слой из PDF-документов}
ddds_prefetch_count = {Максимальное количество одновременно обрабатываемых запросов}
ddas_elastic_apm_server_urls = ${elastic_apm_server_urls}
ddas_elastic_apm_transaction_sample_rate = ${elastic_apm_transaction_sample_rate}
ddas_elastic_apm_span_frames_min_duration =
${elastic_apm_span_frames_min_duration}
ddas_elastic_apm_metrics_interval = ${elastic_apm_metric_interval}
ddas_elastic_apm_capture_body = ${elastic_apm_capture_body}
ddas_elastic_apm_capture_body_content_types = "*"
ddas_queues_message_ttl = ${queues_message_ttl}
ddas_queue_connection = ${rabbit_url}
ddds_queue_retry_connection_count = {Количество повторных попыток подключения к RabbitMQ}
```

- **ddas_replicas.** Количество запускаемых копий сервиса Directum Document Assembler Service (DDAS). Возможные значения: **0** – сервис DDAS не будет запущен при следующем запуске сервисов, **1** – сервис DDAS будет запущен при следующем запуске сервисов. Значение по умолчанию **1**;

- **ddas_external_port.** TCP-порт сервиса Directum Document Assembler Service для обращения к контейнеру через внешнее подключение к хосту. Значение по умолчанию **9057**;
- **ddas_serilog_default.** Параметры ведения лог-файлов.
Подробнее о значениях параметров см. на сервисе GitHub в [документации](#) библиотеки Serilog;
- **ddas_serilog_path_format.** Путь до каталога с лог-файлами Directum Document Assembler Service;
- **ddas_ocr_text_from_pdf.** Всегда извлекать текстовый слой из PDF-документов, даже если текстовый слой уже существует. Возможные значения: **True**, **False**. Значение по умолчанию **True**;
- **ddas_prefetch_count.** Максимальное количество одновременно обрабатываемых сервисом запросов. Значение по умолчанию **1**.
ВАЖНО. Значение параметра **PrefetchCount** должно совпадать со значением аналогичного параметра в конфигурационном файле сервиса **Directum Text Extractor Service**. Это необходимо для корректной работы сервисов с метаданными;
- **ddas_elastic_apm_server_urls.** Адрес сервиса Elastic APM Service. Используется значение, указанное для этого параметра в конфигурационном файле .env в [секции Elastic APM Service](#);
- **ddas_elastic_apm_transaction_sample_rate.** Регулярность отправки трейсов в Elastic APM Service. Используется значение, указанное для этого параметра в конфигурационном файле .env в [секции Elastic APM Service](#);
- **ddas_elastic_apm_span_frames_min_duration.** Если при обработке документа возникла ошибка и время обработки превышает значение параметра, в лог-файл записывается не только факт ошибки, но и подробная информация о ней. Используется значение, указанное для этого параметра в конфигурационном файле .env в [секции Elastic APM Service](#);
- **ddas_elastic_apm_metrics_interval.** Интервал сбора метрик о состоянии памяти и CPU. Используется значение, указанное для этого параметра в конфигурационном файле .env в [секции Elastic APM Service](#);
- **ddas_elastic_apm_capture_body.** Сообщения, которые передаются в сервис Elastic APM Service. Используется значение, указанное для этого параметра в конфигурационном файле .env в [секции Elastic APM Service](#);
- **ddd_queues_message_ttl.** Время хранения сообщений об обработке документа в RabbitMQ в секундах. Используется значение, указанное для [этого параметра](#) в конфигурационном файле .env;
- **ddd_queue_connection.** Строка подключения к очереди сообщений RabbitMQ. Используется значение, указанное для параметра **rabbit_url** в конфигурационном файле .env в [секции RabbitMQ](#);
- **ddd_queue_retry_connection_count.** Количество повторных попыток подключения DSS к брокеру сообщений. Значение по умолчанию **5**.

Text Classifier Service

Сервис Text Classifier Service настраивается в одноименной секции конфигурационного файла `.env`.

Структура секции:

```
# TextClassifierService
dtcs_replicas = {Количество запускаемых копий сервиса Text Classifier Service}
dtcs_external_port = {TCP-порт сервиса Text Classifier Service}
dtcs_debug = {Записывать в лог-файл отладочные сообщения}
dtcs_area = {Используемый раздел конфигурационного файла}
dtcs_max_models_in_pool = {Максимальное количество моделей классификации в пуле}
dtcs_model_cache_lifetime = {Время хранения модели классификации в пуле}
dtcs_dbname = ${pg_database}
dtcs_dbuser = ${pg_user}
dtcs_dbhost = ${pg_host}
dtcs_dbpass = ${pg_password}
dtcs_dbport = ${pg_port}
dtcs_csv_text_field = {Название поля с текстом для обучения данными в формате CSV}
dtcs_csv_class_field = {Название поля с классами для обучения данными в формате CSV многоклассового классификатора}
dtcs_csv_classes_field = {Название поля с классами для обучения данными в формате CSV многозначного классификатора}
dtcs_csv_external_id_field = {Название поля с внешним идентификатором класса для обучения данными в формате CSV многоклассового классификатора}
dtcs_repository = {Тип репозитория для хранения моделей классификации}
dtcs_can_train_models = {Разрешить использовать сервис как сервис обучения классификаторов}
dtcs_queue_connection = ${rabbit_url}
dtcs_message_ttl_seconds = ${queues_message_ttl}
dtcs_prefetch_count = {Максимальное количество одновременно обрабатываемых запросов}
dtcs_max_text_extraction_requests = {Количество документов для одновременного извлечения текста при обучении классификатора}
dtcs_allow_local_path = {Разрешить обучение из локальной папки}
dtcs_server_url = ${elastic_apm_server_urls}
dtcs_metrics_interval = ${elastic_apm_metric_interval}
dtcs_span_frames_min_duration = ${elastic_apm_span_frames_min_duration}
dtcs_suffix = {Формат названия лог-файлов}
```

- **dtcs_replicas.** Количество запускаемых копий сервиса Text Classifier Service. Возможные значения: **0** – сервис DTCS не будет запущен при следующем запуске сервисов, **1** – сервис DTCS будет запущен при следующем запуске сервисов. Значение по умолчанию **1**;
- **dtcs_external_port.** TCP-порт сервиса Text Classifier Service для обращения к контейнеру через внешнее подключение к хосту. Значение по умолчанию **9007**;
- **dtcs_debug.** Записывать в лог-файл отладочные сообщения. Возможные значения: **True**, **False**. Значение по умолчанию **True**;
- **dtcs_area.** Используемый в данный момент раздел конфигурационного файла. Используется для быстрого переключения между настройками окружения. Пример значения: **develop**;

- **dtcs_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**;
- **dtcs_model_cache_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**;
- **dtcs_dbname.** Имя базы данных PostgreSQL. Подставляется значение, указанное для параметра **pg_database** в конфигурационном файле .env в [секции PostgreSQL](#);
- **dtcs_dbuser.** Имя пользователя для подключения к СУБД PostgreSQL. Подставляется значение, указанное для параметра **pg_user** в конфигурационном файле .env в [секции PostgreSQL](#);
- **dtcs_dbhost.** Адрес СУБД PostgreSQL. Подставляется значение, указанное для параметра **pg_host** в конфигурационном файле .env в [секции PostgreSQL](#);
- **dtcs_dbpass.** Пароль для подключения к СУБД PostgreSQL. Подставляется значение, указанное для параметра **pg_password** в конфигурационном файле .env в [секции PostgreSQL](#);
- **dtcs_dbport.** Порт для подключения к СУБД PostgreSQL. Подставляется значение, указанное для параметра **pg_port** в конфигурационном файле .env в [секции PostgreSQL](#);
ПРИМЕЧАНИЕ. Все параметры подключения к базе данных должны совпадать с параметрами, указанными в [конфигурационном файле Smart Service](#);
- **dtcs_csv_text_field, dtcs_csv_class_field, dtcs_csv_classes_field, dtcs_csv_external_id_field.** Названия полей в CSV-файле для обучения классификаторов через набор данных в формате CSV. Параметры являются внутренними, изменять не рекомендуется;
- **dtcs_repository.** Тип репозитория, используемого для хранения моделей классификации. Параметр является внутренним, изменять не рекомендуется;
- **dtcs_can_train_models.** Разрешить использовать этот сервис в качестве сервиса обучения классификаторов. Возможные значения: **True, False**. Значение по умолчанию **True**;
- **dtcs_queue_connection.** Строка подключения к RabbitMQ. Подставляется значение, указанное для параметра **rabbit_url** в конфигурационном файле .env в [секции RabbitMQ](#);
- **dtcs_message_ttl_seconds.** Время хранения сообщений об обработке документа в RabbitMQ в секундах. Подставляется значение, указанное для [данного параметра](#) в конфигурационном файле .env;
- **dtcs_prefetch_count.** Количество одновременно обрабатываемых запросов на сервисе. Значение по умолчанию **0**. Если установлено такое значение, то оптимальное количество одновременно обрабатываемых запросов подбирается автоматически. Оно зависит от доступного ресурса системы, который рассчитывается как количество логических ядер процессора минус одно ядро для работы ОС. Формула для расчета параметра: $(\text{CPU}-1)/2$ с округлением вверх. Например, при CPU=8 доступный ресурс системы будет равен 7 (одно ядро отводится на работу ОС), а итоговое количество параллельно обрабатываемых запросов будет равно 4;
- **dtcs_allow_local_path.** Разрешить обучение классификаторов из локального каталога сервера. Возможные значения: **True, False**. Если установлено значение **False**, повышается безопасность работы. Значение по умолчанию **True**;

- **dtcs_max_text_extraction_requests.** Количество документов без текстового слоя, из которых при обучении классификатора может одновременно извлекаться текст. Рекомендуется указывать значение выше, чем общее количество обработчиков Text Extractor Service, но не слишком большое, так как возрастает нагрузка на CPU. Значение по умолчанию **8**;
- **dtcs_server_url.** Адрес сервиса Elastic APM Service. Подставляется значение, указанное для данного параметра в конфигурационном файле .env в [секции Elastic APM Service](#);
- **dtcs_metrics_interval.** Интервал сбора метрик о состоянии памяти и CPU. Подставляется значение, указанное для данного параметра в конфигурационном файле .env в [секции Elastic APM Service](#);
- **dtcs_span_frames_min_duration.** Сообщения, которые передаются в сервис Elastic APM Service. Подставляется значение, указанное для данного параметра в конфигурационном файле .env в [секции Elastic APM Service](#);
- **dtcs_suffix.** Формат названия лог-файлов. Значение по умолчанию **%Y%m%d**.

Image Classifier Service

Сервис Image Classifier Service настраивается в одноименной секции конфигурационного файла .env.

Структура секции:

```
# ImageClassifierService
dics_replicas = {Количество запускаемых копий сервиса Image Classifier Service}
dics_external_port = {TCP-порт сервиса Image Classifier Service}
dics_debug = {Записывать в лог-файл отладочные сообщения}
dics_area = {Используемый раздел конфигурационного файла}
dics_queue_connection = ${rabbit_url}
dics_message_ttl_seconds = ${queues_message_ttl}
dics_prefetch_count = {Максимальное количество одновременно обрабатываемых запросов}
dics_multithreaded_predict = {Использовать многопоточность}
dics_server_url = ${elastic_apm_server_urls}
dics_metrics_interval = ${elastic_apm_metric_interval}
dics_span_frames_min_duration = ${elastic_apm_span_frames_min_duration}
dics_suffix = {Формат названия лог-файлов}
```

- **dics_replicas.** Количество запускаемых копий сервиса Image Classifier Service. Возможные значения: **0** – сервис DICS не будет запущен при следующем запуске сервисов, **1** – сервис DICS будет запущен при следующем запуске сервисов. Значение по умолчанию **1**;
- **dics_external_port.** TCP-порт сервиса Image Classifier Service для обращения к контейнеру через внешнее подключение к хосту. Значение по умолчанию **9008**;
- **dics_debug.** Записывать в лог-файл отладочные сообщения. Возможные значения: **True**, **False**. Значение по умолчанию **True**;
- **dics_area.** Используемый в данный момент раздел конфигурационного файла. Используется для быстрого переключения между настройками окружения. Пример значения: **develop**;
- **dics_queue_connection.** Строка подключения к RabbitMQ. Подставляется значение, указанное для параметра **rabbit_url** в конфигурационном файле .env в [секции RabbitMQ](#);

- **dics_message_ttl_seconds.** Время хранения сообщений об обработке документа в RabbitMQ в секундах. Подставляется значение, указанное для [данного параметра](#) в конфигурационном файле .env;
- **dics_prefetch_count.** Количество одновременно обрабатываемых запросов на сервисе. Значение по умолчанию **0**. Если установлено такое значение, то оптимальное количество одновременно обрабатываемых запросов подбирается автоматически. Оно зависит от доступного ресурса системы, который рассчитывается как количество логических ядер процессора минус одно ядро для работы ОС. Формула для расчета параметра: **(CPU-1)/2** с округлением вверх. Например, при CPU=8 доступный ресурс системы будет равен 7 (одно ядро отводится на работу ОС), а итоговое количество параллельно обрабатываемых запросов будет равно 4;
- **dics_multithreaded_predict.** Использовать несколько потоков при классификации изображения. Если сервис установлен на отдельном компьютере, установите значение **True**. Значение по умолчанию **False**;
- **dics_server_url.** Адрес сервиса Elastic APM Service. Подставляется значение, указанное для данного параметра в конфигурационном файле .env в [секции Elastic APM Service](#);
- **dics_metrics_interval.** Интервал сбора метрик о состоянии памяти и CPU. Подставляется значение, указанное для данного параметра в конфигурационном файле .env в [секции Elastic APM Service](#);
- **dics_span_frames_min_duration.** Если при обработке документа возникла ошибка и время обработки превышает значение параметра, в лог-файл записывается не только факт ошибки, но и подробная информация о ней. Используется значение, указанное для этого параметра в конфигурационном файле .env в секции Elastic APM Service;
- **dics_suffix.** Формат названия лог-файлов. Значение по умолчанию **%Y%m%d**.

Fact Extractor Base Service

Сервис Fact Extractor Base Service настраивается в одноименной секции конфигурационного файла .env.

Структура секции:

```
# FactExtractorService
dfes_replicas = {Количество запускаемых копий сервиса Fact Extractor Service}
dfes_external_port = {TCP-порт сервиса Fact Extractor Service}
dfes_debug = {Записывать в лог-файл отладочные сообщения}
dfes_area = {Используемый раздел конфигурационного файла}
dfes_queue_connection = ${rabbit_url}
dfes_message_ttl_seconds = ${queues_message_ttl}
dfes_prefetch_count = {Максимальное количество одновременно обрабатываемых запросов}
dfes_server_url = ${elastic_apm_server_urls}
dfes_metrics_interval = ${elastic_apm_metric_interval}
dfes_span_frames_min_duration = ${elastic_apm_span_frames_min_duration}
dfes_suffix = {Формат названия лог-файлов}
```

- **dfes_replicas**. Количество запускаемых копий сервиса Fact Extractor Base Service. Возможные значения: **0** – сервис DFES не будет запущен при следующем запуске сервисов, **1** – сервис DFES будет запущен при следующем запуске сервисов. Значение по умолчанию **1**;
- **dfes_external_port**. TCP-порт сервиса Fact Extractor Base Service для обращения к контейнеру через внешнее подключение к хосту. Значение по умолчанию **9017**;
- **dfes_debug**. Записывать в лог-файл отладочные сообщения. Возможные значения: **True**, **False**. Значение по умолчанию **True**;
- **dfes_area**. Используемый в данный момент раздел конфигурационного файла. Используется для быстрого переключения между настройками окружения. Пример значения: **develop**;
- **dfes_queue_connection**. Строка подключения к RabbitMQ. Подставляется значение, указанное для параметра **rabbit_url** в конфигурационном файле .env в [секции RabbitMQ](#);
- **dfes_message_ttl_seconds**. Время хранения сообщений об обработке документа в RabbitMQ в секундах. Подставляется значение, указанное для [данного параметра](#) в конфигурационном файле .env;
- **dfes_prefetch_count**. Количество одновременно обрабатываемых запросов на сервисе. Значение по умолчанию **0**. Если установлено такое значение, то оптимальное количество одновременно обрабатываемых запросов подбирается автоматически. Оно зависит от доступного ресурса системы, который рассчитывается как количество логических ядер процессора минус одно ядро для работы ОС. Формула для расчета параметра: **(CPU-1)/2** с округлением вверх. Например, при CPU=8 доступный ресурс системы будет равен 7 (одно ядро отводится на работу ОС), а итоговое количество параллельно обрабатываемых запросов будет равно 4.

ПРИМЕЧАНИЕ. Количество одновременно обрабатываемых запросов на экспорт или импорт прикладной разработки всегда равно 1, вне зависимости от установленного в параметре значения;

- **dfes_server_url**. Адрес сервиса Elastic APM Service. Подставляется значение, указанное для данного параметра в конфигурационном файле .env в [секции Elastic APM Service](#);
- **dfes_metrics_interval**. Интервал сбора метрик о состоянии памяти и CPU. Подставляется значение, указанное для данного параметра в конфигурационном файле .env в [секции Elastic APM Service](#);
- **dfes_span_frames_min_duration**. Если при обработке документа возникла ошибка и время обработки превышает значение параметра, в лог-файл записывается не только факт ошибки, но и подробная информация о ней. Используется значение, указанное для этого параметра в конфигурационном файле .env в секции Elastic APM Service;
- **dfes_suffix**. Формат названия лог-файлов. Значение по умолчанию **%Y%m%d**.

Fact Extractor Rules Service

Сервис Fact Extractor Rules Service настраивается в одноименной секции конфигурационного файла `.env`.

Структура секции:

```
# FactExtractorRulesService
dfesr_replicas = {Количество запускаемых копий сервиса Fact Extractor Rules Service}
dfesr_external_port = {TCP-порт сервиса Fact Extractor Rules Service}
dfesr_debug = {Записывать в лог-файл отладочные сообщения}
dfesr_area = {Используемый раздел конфигурационного файла}
dfesr_parser_file_name = {Имя файла синтаксического анализатора для извлечения фактов}
dfesr_parser_debug = {Записывать в лог-файл отладочные сообщения в процессе извлечения фактов и расчета вероятностей}
dfesr_parsers_count = {Количество экземпляров синтаксического анализатора в пуле}
dfesr_parser_timeout = {Таймаут в секундах, после которого обработка запроса прерывается}
dfesr_queue_connection = ${rabbit_url}///?heartbeat={Значение}
dfesr_message_ttl_seconds = ${queues_message_ttl}
dfesr_prefetch_count = {Количество одновременно обрабатываемых сообщений}
dfesr_server_url = ${elastic_apm_server_urls}
dfesr_metrics_interval = ${elastic_apm_metric_interval}
dfesr_span_frames_min_duration = ${elastic_apm_span_frames_min_duration}
dfesr_suffix = {Формат названия лог-файлов}
```

- **dfesr_replicas**. Количество запускаемых копий сервиса Fact Extractor Rules Service. Возможные значения: **0** – сервис DFES-R не будет запущен при следующем запуске сервисов, **1** – сервис DFES-R будет запущен при следующем запуске сервисов. Значение по умолчанию **1**;
- **dfesr_external_port**. TCP-порт сервиса Fact Extractor Rules Service для обращения к контейнеру через внешнее подключение к хосту. Значение по умолчанию **9018**;
- **dfesr_debug**. Записывать в лог-файл отладочные сообщения. Возможные значения: **True**, **False**. Значение по умолчанию **True**;
- **dfesr_area**. Используемый в данный момент раздел конфигурационного файла. Используется для быстрого переключения между настройками окружения. Пример значения: **develop**;
- **dfesr_parser_file_name**. Имя файла синтаксического анализатора, который применяется для извлечения фактов. Значение по умолчанию **parser.exe**. Параметр является внутренним, изменять не рекомендуется;
- **dfesr_parser_debug**. Записывать в лог-файл отладочные сообщения, возникающие в процессе извлечения фактов и расчета вероятностей. Возможные значения: **True**, **False**. Значение по умолчанию **False**;
- **dfesr_parsers_count**. Количество экземпляров синтаксического анализатора, которые хранятся в пуле. Используется для ускорения извлечения фактов. Рекомендуемое значение: на единицу больше, чем количество аппаратных потоков. Значение по умолчанию **4**;

- **dfesr_parser_timeout.** Время, после которого обработка запроса прерывается. В таком случае список извлеченных фактов возвращается пустой. Значение по умолчанию **30s**;
- **dfesr_queue_connection.** Строка подключения к RabbitMQ в формате amqp://{Имя пользователя}:{Пароль}@{Адрес}:{Порт}/{Виртуальный хост}?heartbeat={Значение}, где **heartbeat** – промежуток времени в секундах, за который в RabbitMQ дважды отправляется сообщение о состоянии сервиса. Если этого не происходит, то соединение между RabbitMQ и сервисом прерывается. Значение по умолчанию **300**;
- **dfesr_message_ttl_seconds.** Время хранения сообщений об обработке документа в RabbitMQ в секундах. Подставляется значение, указанное для [данного параметра](#) в конфигурационном файле .env;
- **dfesr_prefetch_count.** Количество одновременно обрабатываемых запросов на сервисе. Значение по умолчанию **0**. Если установлено такое значение, то оптимальное количество одновременно обрабатываемых запросов подбирается автоматически. Оно зависит от доступного ресурса системы, который рассчитывается как количество логических ядер процессора минус одно ядро для работы ОС. Формула для расчета параметра: **(CPU-1)/2** с округлением вверх. Например, при CPU=8 доступный ресурс системы будет равен 7 (одно ядро отводится на работу ОС), а итоговое количество параллельно обрабатываемых запросов будет равно 4.
 ПРИМЕЧАНИЕ. Количество одновременно обрабатываемых запросов на экспорт или импорт прикладной разработки всегда равно 1, вне зависимости от установленного в параметре значения;
- **dfesr_server_url.** Адрес сервиса Elastic APM Service. Подставляется значение, указанное для данного параметра в конфигурационном файле .env в [секции Elastic APM Service](#);
- **dfesr_metrics_interval.** Интервал сбора метрик о состоянии памяти и CPU. Подставляется значение, указанное для данного параметра в конфигурационном файле .env в [секции Elastic APM Service](#);
- **dfesr_span_frames_min_duration.** Если при обработке документа возникла ошибка и время обработки превышает значение параметра, в лог-файл записывается не только факт ошибки, но и подробная информация о ней. Используется значение, указанное для этого параметра в конфигурационном файле .env в секции Elastic APM Service;
- **dfesr_suffix.** Формат названия лог-файла. Значение по умолчанию **%Y%m%d**.

Fact Extractor Learn Service

Сервис Fact Extractor Learn Service настраивается в одноименной секции конфигурационного файла .env.

Структура секции:

```
# FactExtractorLearnService
dfesl_replicas = {Количество запускаемых копий сервиса Fact Extractor Learn Service}
dfesl_max_replicas_per_node = {Ограничение количества запускаемых копий сервиса Fact Extractor Learn Service на одном сервере}
dfesl_external_port = {TCP-порт сервиса Fact Extractor Learn Service}
dfesl_debug = {Записывать в лог-файл отладочные сообщения}
dfesl_worker_cache_lifetime = {Время хранения результата обучения в кэше сервиса в секундах}
```

```

dfesl_can_train_models = {Разрешить использовать сервис как сервис обучения}
dfesl_dbname = ${pg_database}
dfesl_dbuser = ${pg_user}
dfesl_dbhost = ${pg_host}
dfesl_dbpass = ${pg_password}
dfesl_dbport = ${pg_port}
dfesl_queue_connection = ${rabbit_url}///?heartbeat={Значение}
dfesl_message_ttl_seconds = ${queues_message_ttl}
dfesl_prefetch_count = {Количество одновременно обрабатываемых сообщений}
dfesl_model_cache_lifetime = {Время хранения модели в пуле}
dfesl_subprocesses = {Количество подпроцессов}
dfesl_algorithm = {Алгоритм обучения}
dfesl_c1 = {Коэффициент L1-регуляризации}
dfesl_c2 = {Коэффициент L2-регуляризации}
dfesl_epsilon = {Точность вычисления показателя качества модели}
dfesl_delta = {Относительный показатель качества модели}
dfesl_slice_period = {Количество итераций обучения, через которое оценивается
относительный показатель качества модели}
dfesl_max_iterations = {Максимальное число итераций обучения}
dfesl_all_possible_transitions = {Генерировать все сочетания классов}
dfesl_logging_period = {Интервал логирования процесса обучения в итерациях}
dfesl_test_size = {Размер выборки для расчета метрик}
dfesl_server_url = ${elastic_apm_server_urls}
dfesl_metrics_interval = ${elastic_apm_metric_interval}
dfesl_span_frames_min_duration = ${elastic_apm_span_frames_min_duration}
dfesl_suffix = {Формат названия лог-файлов}

```

- **dfesl_replicas.** Количество запускаемых копий сервиса Fact Extractor Learn Service. Возможные значения: **0** – сервис DFES-L не будет запущен при следующем запуске сервисов, **1** – сервис DFES-L будет запущен при следующем запуске сервисов. Значение по умолчанию **1**;
- **dfesl_external_port.** TCP-порт сервиса Fact Extractor Learn Service для обращения к контейнеру через внешнее подключение к хосту. Значение по умолчанию **9019**;
- **dfesl_debug.** Записывать в лог-файл отладочные сообщения. Возможные значения: **True**, **False**. Значение по умолчанию **True**;
- **dfesl_worker_cache_lifetime.** Время хранения результатов обучения модели в кэше сервиса Fact Extractor Learn Service. Указывается в секундах. Если при отправке сообщения о результатах сервису не удастся подключиться к RabbitMQ, то данные о последнем обучении сохраняются в кэше DFES-L в течение заданного периода. Это позволит не запускать обучение заново при восстановлении подключения к RabbitMQ. Значение по умолчанию **86400** (24 часа);
- **dfesl_can_train_models.** Разрешить использовать Fact Extractor Learn Service в качестве сервиса обучения. Возможные значения: **True**, **False**. Значение по умолчанию **True**;
- **dfesl_dbname.** Имя базы данных PostgreSQL. Подставляется значение, указанное для параметра **pg_database** в конфигурационном файле .env в [секции PostgreSQL](#);
- **dfesl_dbuser.** Имя пользователя для подключения к СУБД PostgreSQL. Подставляется значение, указанное для параметра **pg_user** в конфигурационном файле .env в [секции PostgreSQL](#);
- **dfesl_dbhost.** Адрес СУБД PostgreSQL. Подставляется значение, указанное для параметра **pg_host** в конфигурационном файле .env в [секции PostgreSQL](#);

- **dfesi_dbpass.** Пароль для подключения к СУБД PostgreSQL. Подставляется значение, указанное для параметра **pg_password** в конфигурационном файле .env в [секции PostgreSQL](#);
- **dfesi_dbport.** Порт для подключения к СУБД PostgreSQL. Подставляется значение, указанное для параметра **pg_port** в конфигурационном файле .env в [секции PostgreSQL](#);
ПРИМЕЧАНИЕ. Параметры подключения к базе данных должны совпадать с параметрами, указанными в [конфигурационном файле Smart Service](#);
- **dfesi_queue_connection.** Строка подключения к RabbitMQ в формате amqp://{Имя пользователя}:{Пароль}@{Адрес}:{Порт}/{Виртуальный хост}?heartbeat={Значение}, где **heartbeat** – промежуток времени в секундах, за который в RabbitMQ дважды отправляется сообщение о состоянии сервиса. Если этого не происходит, то соединение между RabbitMQ и сервисом прерывается. Значение по умолчанию **36000**;
- **dfesi_message_ttl_seconds.** Время хранения сообщений об обработке документа в RabbitMQ в секундах. Значение по умолчанию **86400**. Подставляется значение, указанное для [данного параметра](#) в конфигурационном файле .env;
- **dfesi_prefetch_count.** Количество одновременно обрабатываемых запросов на сервисе. Значение по умолчанию **0**. Если установлено такое значение, то оптимальное количество одновременно обрабатываемых запросов подбирается автоматически. Оно зависит от доступного ресурса системы, который рассчитывается как количество логических ядер процессора минус одно ядро для работы ОС. Формула для расчета параметра: $(CPU-1)/2$ с округлением вверх. Например, при CPU=8 доступный ресурс системы будет равен 7 (одно ядро отводится на работу ОС), а итоговое количество параллельно обрабатываемых запросов будет равно 4.

ПРИМЕЧАНИЕ. На некоторые типы запросов распространяются дополнительные условия, которые позволяют снизить нагрузку на систему. А именно:

- если значение параметра **0**, то количество одновременно обрабатываемых запросов на обучение или дообучение моделей всегда равно 1;
- количество одновременно обрабатываемых запросов на экспорт или импорт прикладной разработки всегда равно 1, вне зависимости от установленного в параметре значения;
- **dfesi_model_cache_lifetime.** Время хранения модели в невыгружаемом пуле в минутах. Модели, к которым не обращаются в течение этого времени, выгружаются из пула. Укажите целое число больше нуля. Значение по умолчанию **60**;
- **dfesi_subprocesses.** Количество подпроцессов для более быстрого обучения моделей извлечения фактов. Если параметр отсутствует, подпроцессы не выделяются. Значение по умолчанию **4**;
- **dfesi_algorithm.** Алгоритм обучения модели. Поддерживается только алгоритм lbfgs;
- **dfesi_c1, dfesi_c2.** Значение коэффициентов по умолчанию **0,1**;
- **dfesi_epsilon.** Точность, с которой вычисляется параметр **dfesi_delta**. Значение по умолчанию **0,00001**;
- **dfesi_delta.** Допустимая разница в качестве обучения модели на разных итерациях. Если качество модели отличается не более чем на установленное значение, то считается, что качество не изменилось. Значение по умолчанию **0,04**;
- **dfesi_slice_period.** Количество итераций обучения, через которое проверяется, что значение параметра **delta** не превышает установленного. Значение по умолчанию **10**;

- **dfesl_max_iterations.** Максимальное количество итераций для обучения. Значение по умолчанию **250**;
- **dfesl_all_possible_transitions.** Значение по умолчанию **True**;
- **dfesl_logging_period.** Периодичность логирования процесса при обучении модели извлечения фактов. Задается в итерациях. Значение по умолчанию **10**. Если логирование процесса обучения не требуется, отключите его, указав значение **0**;
- **dfesl_test_size.** Значение по умолчанию **0,25**.
ПРИМЕЧАНИЕ. Подробнее о параметрах обучения см. в документации к библиотеке Sklearn-crfsuite статью [API Reference](#), раздел CRF;
- **dfesl_server_url.** Адрес сервиса Elastic APM Service. Подставляется значение, указанное для данного параметра в конфигурационном файле .env в [секции Elastic APM Service](#);
- **dfesl_metrics_interval.** Интервал сбора метрик о состоянии памяти и CPU. Подставляется значение, указанное для данного параметра в конфигурационном файле .env в [секции Elastic APM Service](#);
- **dfesl_span_frames_min_duration.** Если при обработке документа возникла ошибка и время обработки превышает значение параметра, в лог-файл записывается не только факт ошибки, но и подробная информация о ней. Используется значение, указанное для этого параметра в конфигурационном файле .env в секции Elastic APM Service;
- **dfesl_suffix.** Формат названия лог-файла. Значение по умолчанию **%Y%m%d**.

Object Detection Service

Сервис Object Detection Service настраивается в одноименной секции конфигурационного файла .env.

Структура секции:

```
# ObjectDetectionService
dods_replicas = {Количество запускаемых копий сервиса Object Detection Service}
dods_external_port = {TCP-порт сервиса Fact Extractor Learn Service}
dods_debug = {Записывать в лог-файл отладочные сообщения}
dods_queue_connection = ${rabbit_url}
dods_message_ttl_seconds = ${queues_message_ttl}
dods_prefetch_count = {Максимальное количество одновременно обрабатываемых запросов}
dods_server_url = ${elastic_apm_server_urls}
dods_metrics_interval = ${elastic_apm_metric_interval}
dods_span_frames_min_duration = ${elastic_apm_span_frames_min_duration}
dods_suffix = {Формат названия лог-файлов}
dods_debug_images = {Режим отладки для сохранения промежуточных результатов обработки изображений}
dods_output_directory = {Каталог для сохранения промежуточных результатов обработки изображений внутри контейнера}
dods_output_directory_host = {Каталог для сохранения промежуточных результатов обработки изображений}
```

- **dods_replicas.** Количество запускаемых копий сервиса Object Detection Service. Возможные значения: **0** – сервис DODS не будет запущен при следующем запуске сервисов, **1** – сервис DODS будет запущен при следующем запуске сервисов. Значение по умолчанию **1**;
- **dods_external_port.** TCP-порт сервиса Object Detection Service для обращения к контейнеру через внешнее подключение к хосту. Значение по умолчанию **9047**;

- **dods_debug.** Записывать в лог-файл отладочные сообщения. Возможные значения: **True**, **False**. Значение по умолчанию **True**;
- **dods_queue_connection.** Строка подключения к RabbitMQ. Подставляется значение, указанное для параметра **rabbit_url** в конфигурационном файле `.env` в [секции RabbitMQ](#);
- **dods_message_ttl_seconds.** Время хранения сообщений об обработке документа в RabbitMQ в секундах. Подставляется значение, указанное для [данного параметра](#) в конфигурационном файле `.env`;
- **dods_prefetch_count.** Количество одновременно обрабатываемых запросов на сервисе. Значение по умолчанию **0**. Если установлено такое значение, то оптимальное количество одновременно обрабатываемых запросов подбирается автоматически. Оно зависит от доступного ресурса системы, который рассчитывается как количество логических ядер процессора минус одно ядро для работы ОС. Формула для расчета параметра: $(CPU-1)/2$ с округлением вверх. Например, при CPU=8 доступный ресурс системы будет равен 7 (одно ядро отводится на работу ОС), а итоговое количество параллельно обрабатываемых запросов будет равно 4;
- **dods_server_url.** Адрес сервиса Elastic APM Service. Подставляется значение, указанное для данного параметра в конфигурационном файле `.env` в [секции Elastic APM Service](#);
- **dods_metrics_interval.** Интервал сбора метрик о состоянии памяти и CPU. Подставляется значение, указанное для данного параметра в конфигурационном файле `.env` в [секции Elastic APM Service](#);
- **dods_span_frames_min_duration.** Если при обработке документа возникла ошибка и время обработки превышает значение параметра, в лог-файл записывается не только факт ошибки, но и подробная информация о ней. Используется значение, указанное для этого параметра в конфигурационном файле `.env` в секции Elastic APM Service;
- **dods_suffix.** Формат названия лог-файлов. Значение по умолчанию **%Y%m%d**;
- **dods_debug_images.** Режим отладки для сохранения промежуточных результатов обработки изображений. Возможные значения: **True**, **False**. Если указано значение **True**, то во временный каталог, указанный в параметре **dods_output_directory_host**, сохраняются промежуточные результаты обработки изображений на каждом из этапов. Значение по умолчанию **False**.

Чтобы включить режим отладки:

1. В параметре **dods_debug_images** установите значение **True**.

2. В конфигурационном файле `deploy.yml` в секции **dods** в разделе **volumes** раскомментируйте строку:

```
#- ${dods_output_directory_host}:/usr/src/objectdetection-
service/${dods_output_directory}
```

3. Перезапустите сервисы Ario. Для этого последовательно выполните команды:

```
./do.sh ario down
./do.sh ario up
```

- **dods_output_directory.** Адрес временного каталога для сохранения промежуточных результатов обработки изображений. Значение по умолчанию **image_masks**. Не рекомендуется менять значение параметра;
- **dods_output_directory_host.** Каталог для сохранения промежуточных результатов обработки изображений на сервере. Значение по умолчанию **./etc/objectdetection-service/image_masks**. Не рекомендуется менять значение параметра.

Storage Service

Сервис Storage Service настраивается в одноименной секции конфигурационного файла `.env`.

Структура секции:

```
# StorageService
storage_service_replicas = {Количество запускаемых копий сервиса хранения объектов}
storage_service_external_port = {TCP-порт сервиса хранения объектов}
storage_service_access_key = {Имя пользователя сервиса хранения объектов}
storage_service_secret_key = {Пароль пользователя сервиса хранения объектов}
```

- **storage_service_replicas.** Количество запускаемых копий сервиса хранения объектов. Возможные значения: **0** – сервис хранения объектов не будет запущен при следующем запуске сервисов, **1** – сервис хранения объектов будет запущен при следующем запуске сервисов. Значение по умолчанию **1**;
- **storage_service_external_port.** TCP-порт сервиса хранения объектов для обращения к контейнеру через внешнее подключение к хосту. Значение по умолчанию **9021**;
- **storage_service_access_key.** Имя пользователя для доступа к хранилищу;
- **storage_service_secret_key.** Пароль пользователя для доступа к хранилищу.

Elastic APM Service

Сервис Elastic APM Service настраивается в одноименной секции конфигурационного файла `.env`.

Структура секции:

```
# ElasticApm
elastic_apm_replicas = {Количество запускаемых копий сервиса Elastic APM Service}
elastic_apm_external_port = {TCP-порт сервиса Elastic APM Service}
elastic_apm_server_urls = {Адрес сервиса Elastic APM Service}
elastic_apm_log_level = {Минимальный уровень логирования сервиса Elastic APM Service}
elastic_apm_transaction_sample_rate = {Регулярность отправки трейсов в сервис Elastic APM Service}
elastic_apm_span_frames_min_duration = {Время, в течение которого в лог файл записывается подробная информация об ошибке обработки документа. Указывается в миллисекундах}
elastic_apm_metric_interval = {Интервал сбора метрик в секундах}
elastic_apm_capture_body = {Сообщения, которые передаются в сервис Elastic APM Service}
elastic_apm_traces_path = {Путь до каталога с файлами трассировок}
elastic_apm_traces_file_name = {Базовое имя файлов трассировки}
elastic_apm_output_file_rotate_every_kb = {Максимальный размер файла трассировки}
elastic_apm_output_file_number_of_files = {Максимальное количество файлов трассировки}
```

- **elastic_apm_replicas**. Количество запускаемых копий сервиса Elastic APM Service. Возможные значения: **0** – сервис Elastic APM Service не будет запущен при следующем запуске сервисов, **1** – сервис Elastic APM Service будет запущен при следующем запуске сервисов. Значение по умолчанию **1**;
- **elastic_apm_external_port**. TCP-порт сервиса Elastic APM Service для обращения к контейнеру через внешнее подключение к хосту. Значение по умолчанию **9082**;
- **elastic_apm_server_urls**. Адрес сервиса Elastic APM Service;
- **elastic_apm_log_level**. Минимальный уровень протоколирования Elastic APM Service;
- **elastic_apm_transaction_sample_rate**. Регулярность отправки трейсов в Elastic APM Service. Укажите значение **1.0**, если необходимо отправлять все трейсы, **0.5** – если каждый второй. Значение по умолчанию **1.0**;
- **elastic_apm_span_frames_min_duration**. Если при обработке документа возникла ошибка и время обработки превышает значение параметра, в лог-файл записывается не только факт ошибки, но и подробная информация о ней. Значение по умолчанию **0ms**, при котором подробная запись информации об ошибке в лог-файл отключена;
- **elastic_apm_metric_interval**. Интервал сбора метрик о состоянии памяти и CPU. Значение по умолчанию **0s**, при котором метрики не собираются;
- **elastic_apm_capture_body**. Сообщения, которые передаются в сервис Elastic APM Service. Возможные значения:
 - **off** – сообщения не передаются;
 - **errors** – сообщения с ошибками;
 - **transactions** – сообщения без ошибок;
 - **all** – все сообщения.
 Значение по умолчанию **all**;
- **elastic_apm_traces_path**. Путь до каталога с файлами трассировок. Значение по умолчанию **./etc/elastic-apm/traces**;
- **elastic_apm_file_name**. Базовое имя файлов трассировок. Например, при значении **apm-server** файлам присваиваются имена **apm-server**, **apm-server.1**, **apm-server.2** и т.д.;
- **elastic_apm_output_file_rotate_every_kb**. Максимальный размер файла в килобайтах. Когда файл с трейсами достигает максимального размера, создается новый файл. Значение по умолчанию **10240**;
- **elastic_apm_output_file_number_of_files**. Максимальное количество файлов трассировок. Если количество файлов превышено, старые файлы перезаписываются. Значение по умолчанию **7**.

Portainer

Параметры сервиса Portainer, который предоставляет графический веб-интерфейс для управления контейнерами после их запуска, настраиваются в одноименной секции конфигурационного файла `.env`. Portainer используется вместе с Portainer Agent.

Структура секции:

```
# Portainer
portainer_external_port = {TCP-порт Portainer}
```

- **portainer_external_port.** TCP-порт сервиса Portainer для обращения к контейнеру через внешнее подключение к хосту. Значение по умолчанию **8999**.

RabbitMQ

Параметры запуска брокера сообщений RabbitMQ настраиваются в одноименной секции конфигурационного файла .env.

Структура секции:

```
# RabbitMQ
rabbit_replicas = {Количество запускаемых копий брокера сообщений}
rabbit_external_port = {TCP-порт брокера сообщений}
rabbit_user = {Имя пользователя брокера сообщений}
rabbit_password = {Пароль пользователя брокера сообщений}
rabbit_url = {Строка подключения к RabbitMQ}
```

- **rabbit_replicas.** Количество запускаемых копий брокера сообщений RabbitMQ. Возможные значения: **0** – брокер сообщений не будет запущен при следующем запуске сервисов, **1** – брокер сообщений будет запущен при следующем запуске сервисов. Значение по умолчанию **1**;
- **rabbit_external_port.** TCP-порт сервиса RabbitMQ для обращения к контейнеру через внешнее подключение к хосту. Значение по умолчанию **5672**;
- **rabbit_user** и **rabbit_password.** Имя и пароль пользователя брокера сообщений;
- **rabbit_url.** Строка подключения к брокеру сообщений RabbitMQ в формате `amqp://{Имя пользователя}:{Пароль}@{Адрес}:{Порт}/{Virtual Host}`, где:
 - **Виртуальный хост** – виртуальный тенант RabbitMQ. Подробнее см. в документации RabbitMQ статью [Virtual Hosts](#).

Пример значения: **amqp://ario:ario12345@localhost:5672/ario**;

PostgreSQL

Параметры запуска СУБД PostgreSQL настраиваются в одноименной секции конфигурационного файла .env.

Структура секции:

```
# PostgreSQL
pg_replicas = {Количество запускаемых копий СУБД PostgreSQL}
pg_external_port = {TCP-порт СУБД PostgreSQL}
pg_host = {Адрес СУБД PostgreSQL}
pg_user = {Имя пользователя для подключения к СУБД PostgreSQL}
pg_password = {Пароль для подключения к СУБД PostgreSQL}
pg_database = {Имя базы данных PostgreSQL для Smart Service}
pg_database_dis = {Имя базы данных PostgreSQL для Identity Service}
pg_port = {Порт для подключения к СУБД PostgreSQL}
pg_connection = User
ID=${pg_user};Password=${pg_password};Host=${pg_host};Port=${pg_port};Database=${pg_database};Pooling={Использовать пул подключений к БД}
```

- **pg_replicas**. Количество запускаемых копий СУБД PostgreSQL. Возможные значения: **0** – СУБД PostgreSQL не будет запущена при следующем запуске сервисов, **1** – СУБД PostgreSQL будет запущена при следующем запуске сервисов. Значение по умолчанию **1**;
- **pg_external_port**. TCP-порт СУБД PostgreSQL для обращения к контейнеру через внешнее подключение к хосту. Значение по умолчанию **5432**;
- **pg_host**. Адрес СУБД PostgreSQL. Значение по умолчанию **localhost**;
- **pg_user** и **pg_password**. Имя и пароль пользователя для подключения к СУБД PostgreSQL;
- **pg_database**. Имя базы данных PostgreSQL для Smart Service;
- **pg_database_dis**. Имя базы данных PostgreSQL для Identity Service;
- **pg_port**. TCP-порт СУБД PostgreSQL для обращения к контейнеру внутри хоста. Значение по умолчанию **5432**;
- **pg_connection**. Строка подключения к СУБД PostgreSQL в формате:
User
ID=\${pg_user};Password=\${pg_password};Host=\${pg_host};Port=\${pg_port};Database=\${pg_database};Pooling={Использовать пул подключений к БД}.

InfluxDB

Параметры базы данных InfluxDB, которая предназначена для хранения данных из Telegraf и их передачи в Grafana, настраиваются в одноименной секции конфигурационного файла .env.

Структура секции:

```
# InfluxDB
influxdb_replicas = {Количество запускаемых копий InfluxDB}
influxdb_user = {Имя пользователя InfluxDB}
influxdb_password = {Пароль пользователя InfluxDB}
```

- **influxdb_replicas**. Количество запускаемых копий базы данных InfluxDB. Возможные значения: **0** – база данных InfluxDB не будет запущена при следующем запуске сервисов, **1** – база данных InfluxDB будет запущена при следующем запуске сервисов. Значение по умолчанию **1**;
- **influxdb_user** и **influxdb_password**. Имя и пароль пользователя InfluxDB. Значение по умолчанию **admin**.

Grafana

Параметры информационной панели Grafana, которая предназначена для отображения метрик в реальном времени, настраиваются в одноименной секции конфигурационного файла .env.

Структура секции:

```
# Grafana
grafana_replicas = {Количество запускаемых копий Grafana}
grafana_external_port = {TCP-порт Grafana}
```

```
grafana_user = {Имя пользователя Grafana}
grafana_password = {Пароль пользователя Grafana}
```

- **grafana_replicas**. Количество запускаемых копий информационной панели Grafana. Возможные значения: **0** – Grafana не будет запущена при следующем запуске сервисов, **1** – Grafana будет запущена при следующем запуске сервисов. Значение по умолчанию **1**;
- **grafana_external_port**. TCP-порт Grafana для обращения к контейнеру через внешнее подключение к хосту. Значение по умолчанию **3000**;
- **grafana_user** и **grafana_password**. Имя и пароль пользователя Grafana. Значение по умолчанию **admin**.

Telegraf

Параметры серверного агента Telegraf, который собирает информацию о метриках и передает ее в базу данных InfluxDB, настраиваются в одноименной секции конфигурационного файла `.env`.

Структура секции:

```
# Telegraf
telegraf_replicas = {Количество запускаемых копий Telegraf}
```

- **telegraf_replicas**. Количество запускаемых копий Telegraf. Возможные значения: **0** – Telegraf не будет запущен при следующем запуске сервисов, **1** – Telegraf будет запущен при следующем запуске сервисов. Значение по умолчанию **1**.

Общие параметры

- **ario_version**. Номер текущей версии сервисов. Пример значения: **2021.12.2.0**;
- **queues_message_ttl**. Время хранения сообщений об обработке документа в RabbitMQ в секундах. Значение по умолчанию **86400**.

Чтобы изменить время хранения сообщений:

1. Запросите в службе поддержки Directum скрипт удаления очередей RabbitMQ и расположите его в корневом каталоге сервисов рядом со скриптами остановки и удаления.
2. Измените значение параметра **queues_message_ttl** в конфигурационном файле `.env`.
3. Запустите скрипт, не останавливая сервисы Ario.

ContainersTimezone – параметры времени внутри контейнеров:

- **timezone**. Часовой пояс внутри контейнеров. Значение по умолчанию **Europe/Samara**.

Logs – параметры логирования сервисов:

- **log_path**. Путь до корневого каталога с лог-файлами. Значение по умолчанию **./logs/**;
- **log_level**. Минимальный уровень логирования. Значение по умолчанию **Information**;

- **log_path_format.** Формат имени лог-файла. Значение по умолчанию **{Date}.log**;
- **dss_logs, dtes_logs, dis_logs, dps_logs, dfes_logs, dfesr_logs, dfesl_logs, dods_logs, dics_logs, dtcs_logs.** Подкаталоги для лог-файлов каждого сервиса Ario в корневом каталоге. Пример значения: **\$(log_path)textextractor-service.**

ServicePlacement – параметры расположения сервисов на серверах. По умолчанию все сервисы запускаются на управляющем сервере, где значение **label** по умолчанию – **ario**:

- **manager_node_label.** В параметре должно дублироваться значение **label** управляющего сервера;
- **pg_node_label.** Содержит значение **label** сервера, на котором развернута СУБД PostgreSQL;
- **rabbit_node_label.** Содержит значение **label** сервера, на котором развернут брокер сообщений RabbitMQ;
- **storage_service_node_label, dtes_node_label, dfes_node_label, dfesr_node_label, dfesl_node_label, dods_node_label, dics_node_label, dtcs_node_label.** Содержит значение **label** сервера, на котором развернуты соответствующие сервисы Ario.

Повторная активация сервисов

Для повторной активации сервисов, например, при истечении срока действия лицензии или изменении конфигурации оборудования:

1. Запросите ключ активации в службе поддержки Directum.
2. Остановите работу сервиса Directum Smart Service. Для этого выполните команду:

```
./do.sh ario_dss down
```

3. Удалите файл LicenseKey.alk из папки с установленным сервисом Directum Smart Service и скопируйте в нее ключ активации ActivationKey.aak.
4. Запустите сервис Directum Smart Service. Для этого выполните команду:

```
./do.sh ario_dss up
```

ПРИМЕЧАНИЕ. При необходимости можно останавливать и запускать все сервисы с помощью команд:

- для остановки сервисов:

```
./do.sh ario down
```

- для запуска сервисов:

```
./do.sh ario up
```

Справочная информация

Ознакомьтесь со справочной информацией:

- [виды документов](#), которые обрабатывают сервисы Агіо;
- [список фактов](#), которые умеют извлекать сервисы Агіо;
- [ограничения](#) при извлечении фактов;
- [список языков](#), доступных для распознавания.

Виды обрабатываемых документов

В базовой поставке Directum Агіо поддерживается обработка стандартных видов документов. Полнота извлечения *фактов* для них составляет 80% и выше.

Для корректной обработки документы следует оформлять по принятым формам:

Вид документа, включенный в базовую поставку	Рекомендация к оформлению документов
Акт выполненных работ	Общепринятая практика делового документооборота
Входящее письмо	ГОСТ Р 7.0.97-2016
Железнодорожный билет (электронный)	Рекомендуемая форма, установленная ОАО «РЖД»
Паспорт гражданина РФ	Установленная форма
Постановление судебных приставов	Приложение N 103 к приказу Федеральной службы судебных приставов N 318
Страховое свидетельство об обязательном пенсионном страховании	Установленная форма
Счет на оплату (русскоязычный)	Общепринятая практика делового документооборота
Счет-фактура	Приложение N 1 к постановлению Правительства РФ N 1137
Товарная накладная	Унифицированная форма № ТОРГ-12
Уведомление о регистрации в системе пенсионного учета (форма АДИ РЕГ)	Приложение 1 к постановлению Правления Пенсионного фонда Российской Федерации № 335п
Универсальный передаточный документ	Рекомендуемая форма

Сервисы Агіо также обрабатывают дополнительные виды документов, для которых полнота извлечения фактов может быть ниже 80%:

- авиабилет;
- договор;
- дополнительное соглашение;
- корректировочный счет-фактура;
- свидетельство о постановке на налоговый учет;
- счет на оплату (англоязычный);
- товарно-транспортная накладная;
- транспортная железнодорожная накладная;

- транспортная накладная;
- универсальный корректировочный документ;
- экспедиторская расписка.

Список извлекаемых фактов

Вид документа	Извлекаемые факты
Входящее письмо	<p>Вид документа</p> <p>Дата документа</p> <p>Номер документа</p> <p>Дата документа «В ответ на»</p> <p>Номер документа «В ответ на»</p> <p>Прочие даты документа</p> <p>Организационно-правовая форма (ОПФ) и Наименование контрагента</p> <p>Наименование головной организации</p> <p>ИНН и КПП контрагента</p> <p>ОГРН и ОГРНИП контрагента</p> <p>Р/с контрагента</p> <p>Наименование и БИК банка контрагента</p> <p>Номер телефона контрагента</p> <p>Адрес электронной почты контрагента</p> <p>Сайт организации контрагента</p> <p>Адресат</p> <p>Подписант письма</p> <p>Исполнитель</p> <p>Тип персоны</p> <p>Тема письма, начинающаяся с «О» или «Об» и расположенная вверху первой страницы письма</p> <p>Гриф доступа: «Конфиденциально», «Для служебного пользования», «Коммерческая тайна»</p> <p>Наименование прочей организации</p> <p>ИНН и КПП прочей организации</p> <p>ОГРН и ОГРНИП прочей организации</p> <p>Р/с прочей организации</p> <p>Наименование и БИК банка прочей организации</p>
Договор (в том числе на английском языке)	<p>Документы на русском языке:</p> <p>Вид документа</p> <p>Номер договора</p> <p>Дата договора</p> <p>Сумма</p> <p>ОПФ и Наименование контрагента</p>

Вид документа	Извлекаемые факты
Дополнительное соглашение	ИНН и КПП контрагента
	Количество листов
	Валюта
	Документы на английском языке:
	Вид документа
	Номер договора
	Дата договора
	Наименование контрагента
	Номер документа
	Дата документа
Номер ведущего документа	
Дата ведущего документа	
ОПФ и Наименование контрагента	
ИНН и КПП контрагента	
Подписант дополнительного соглашения	
Счет-фактура	Вид документа
	Номер документа
	Дата документа
	Номер исправления
	Дата исправления
	Прочие даты документа
	ОПФ и Наименование контрагента
	ИНН и КПП контрагента
	Адрес контрагента
	Тип контрагента (грузоотправитель, грузополучатель, продавец, покупатель)
	Сумма без НДС, по документу
	Сумма НДС, по документу
	Сумма с НДС, по документу
	Валюта
	Наименование товара
	Код вида товара
	Единица измерения
	Код единицы измерения по ОКЕИ
	Количество товара (объем)
	Цена за единицу, по строке
	Сумма без НДС, по строке
	Сумма акциза, по строке
	Ставка НДС, по строке
	Сумма НДС, по строке

Вид документа	Извлекаемые факты
Корректировочный счет-фактура	<p>Сумма с НДС, по строке</p> <p>Страна происхождения товара (код и краткое наименование)</p> <p>Регистрационный номер таможенной декларации</p> <p>Вид документа</p> <p>Номер документа</p> <p>Дата документа</p> <p>Номер исправления корректировки</p> <p>Дата исправления корректировки</p> <p>Номер корректируемого документа</p> <p>Дата корректируемого документа</p> <p>ОПФ и Наименование контрагента</p> <p>ИНН и КПП контрагента</p> <p>Адрес контрагента</p> <p>Тип контрагента (продавец, покупатель)</p> <p>Валюта документа</p>
Уведомление о регистрации в системе пенсионного учета	<p>Страховой номер</p> <p>ФИО</p> <p>Дата рождения</p> <p>Место рождения</p> <p>Пол</p> <p>Дата регистрации в системе</p>
Страховое свидетельство об обязательном пенсионном страховании	<p>Страховой номер</p> <p>ФИО</p> <p>Дата рождения</p> <p>Пол</p>
Универсальный передаточный документ	<p>Вид документа</p> <p>Номер документа</p> <p>Дата документа</p> <p>Номер исправления</p> <p>Дата исправления</p> <p>Наименование документа-основания</p> <p>Дата документа-основания</p> <p>Номер документа-основания</p> <p>Статус (функция) УПД</p> <p>Прочие даты документа</p> <p>ОПФ и Наименование контрагента</p> <p>ИНН и КПП контрагента</p> <p>Адрес контрагента</p> <p>Тип контрагента (грузоотправитель, грузополучатель, продавец, покупатель)</p>

Вид документа	Извлекаемые факты
	<p>Сумма без НДС, по документу</p> <p>Сумма НДС, по документу</p> <p>Сумма с НДС, по документу</p> <p>Количество листов</p> <p>Валюта</p> <p>Код товара</p> <p>Наименование товара</p> <p>Код вида товара</p> <p>Количество товара (объем)</p> <p>Единица измерения</p> <p>Код единицы измерения по ОКЕИ</p> <p>Цена за единицу, по строке</p> <p>Сумма акциза, по строке</p> <p>Сумма без НДС, по строке</p> <p>Ставка НДС, по строке</p> <p>Сумма НДС, по строке</p> <p>Сумма с НДС, по строке</p> <p>Страна происхождения товара (код и краткое наименование)</p> <p>Регистрационный номер таможенной декларации</p>
<p>Универсальный корректировочный документ</p>	<p>Вид документа</p> <p>Номер документа</p> <p>Дата документа</p> <p>Номер исправления корректировки</p> <p>Дата исправления корректировки</p> <p>Номер корректируемого документа</p> <p>Дата корректируемого документа</p> <p>ОПФ и Наименование контрагента</p> <p>ИНН и КПП контрагента</p> <p>Адрес контрагента</p> <p>Тип контрагента (продавец, покупатель)</p> <p>Валюта документа</p> <p>Статус (функция) УКД</p>
<p>Товарная накладная (ТОРГ-12)</p>	<p>Вид документа</p> <p>Номер документа</p> <p>Дата документа</p> <p>Прочие даты документа</p> <p>ОПФ и Наименование контрагента</p> <p>ИНН и КПП контрагента</p> <p>Адрес контрагента</p>

Вид документа	Извлекаемые факты
	<p>Р/с контрагента</p> <p>Наименование и БИК банка контрагента</p> <p>Тип контрагента (грузоотправитель, грузополучатель, поставщик, плательщик)</p> <p>Сумма НДС, по документу</p> <p>Сумма с НДС, по документу</p> <p>Сумма без НДС, по документу</p> <p>Наименование документа-основания</p> <p>Дата документа-основания</p> <p>Номер документа-основания</p> <p>Номер транспортной накладной</p> <p>Дата транспортной накладной</p> <p>Валюта</p> <p>Наименование товара</p> <p>Код товара</p> <p>Единица измерения</p> <p>Код единицы измерения по ОКЕИ</p> <p>Вид упаковки</p> <p>Количество в одном месте</p> <p>Количество мест/штук</p> <p>Масса брутто, по строке</p> <p>Масса нетто, по строке</p> <p>Цена за единицу, по строке</p> <p>Сумма без НДС, по строке</p> <p>Ставка НДС, по строке</p> <p>Сумма НДС, по строке</p> <p>Сумма с НДС, по строке</p>
Акт выполненных работ	<p>Вид документа</p> <p>Номер документа</p> <p>Дата документа</p> <p>Прочие даты документа</p> <p>Наименование документа-основания</p> <p>Дата документа-основания</p> <p>Номер документа-основания</p> <p>Сумма без НДС, по документу</p> <p>Сумма НДС, по документу</p> <p>Сумма с НДС, по документу</p> <p>Валюта</p> <p>Целая часть суммы, выраженная числом</p> <p>Дробная часть суммы, выраженная числом</p>

Вид документа	Извлекаемые факты
	<p>Целая часть суммы, выраженная прописью Дробная часть суммы, выраженная прописью Валюта целой части Валюта дробной части ОПФ и Наименование контрагента ИНН и КПП контрагента ОГРН и ОГРНИП контрагента Р/с контрагента Наименование и БИК банка контрагента ФИО подписанта контрагента Тип контрагента (продавец, покупатель) Наименование работ/услуг Количество работ/услуг Единица измерения Цена за единицу, по строке Сумма с НДС, по строке</p>
Счет на оплату (русскоязычный)	<p>Вид документа Номер документа Дата документа Наименование документа-основания Дата документа-основания Номер документа-основания Номер заказа на поставку Дата получения Прочие даты документа Сумма без НДС, по документу Сумма НДС, по документу Сумма с НДС, по документу Валюта Целая часть суммы, выраженная числом Дробная часть суммы, выраженная числом Целая часть суммы, выраженная прописью Дробная часть суммы, выраженная прописью Валюта целой части Валюта дробной части ОПФ и Наименование контрагента ИНН и КПП контрагента ОГРН и ОГРНИП контрагента Р/с контрагента Наименование и БИК банка контрагента</p>

Вид документа	Извлекаемые факты
	Тип контрагента (продавец, покупатель) Код товара Наименование товара Количество товара Единица измерения Цена за единицу, по строке Ставка НДС, по строке Сумма НДС, по строке Сумма с НДС, по строке
Счет на оплату (англоязычный)	Вид документа Номер документа Дата документа Прочие даты документа Наименование контрагента Тип контрагента (продавец, покупатель) Номер заказа Банковский счет Сумма документа Сумма НДС Наименование документа-основания Дата документа-основания Номер документа-основания Валюта
Транспортная накладная (ТРН)	Вид документа Номер документа Дата документа
Товарно-транспортная накладная (ТТН)	Вид документа Номер документа Дата документа ОПФ и Наименование контрагента ИНН и КПП контрагента Адрес контрагента Тип контрагента (грузоотправитель, грузополучатель, плательщик)
Транспортная железнодорожная накладная	Вид документа Номер документа Дата документа
Экспедиторская расписка	Вид документа Номер документа Дата документа

Вид документа	Извлекаемые факты
Паспорт гражданина РФ	<p>Вид документа</p> <p>Серия и номер паспорта</p> <p>ФИО гражданина</p> <p>Пол гражданина</p> <p>Дата выдачи паспорта</p> <p>Дата рождения</p> <p>Место рождения</p> <p>Наименование организации, выдавшей паспорт</p> <p>Код подразделения организации, выдавшей паспорт</p> <p>Машиночитаемая зона</p>
Свидетельство о постановке на налоговый учет	<p>Вид документа</p> <p>ОПФ и Наименование организации/ФИО физического лица</p> <p>ИНН организации/физического лица</p> <p>КПП (для организаций)</p> <p>ОГРН (для организаций)</p> <p>Дата постановки на налоговый учет</p> <p>Серия и номер документа</p>
Авиабилет (электронный)	<p>Вид документа</p> <p>Маршрут авиабилета</p> <p>Дата отправления</p> <p>Дата прибытия</p> <p>Город отправления</p> <p>Город прибытия</p> <p>Стоимость билета (сумма и код валюты)</p> <p>ФИО персоны</p>
Железнодорожный билет (электронный)	<p>Вид документа</p> <p>Дата отправления</p> <p>Дата прибытия</p> <p>Маршрут</p> <p>Стоимость билета (сумма и код валюты)</p> <p>ФИО персоны</p>
Постановление судебного пристава-исполнителя	<p>Вид документа</p> <p>Номер и дата документа</p> <p>Номер и дата исполнительного производства</p> <p>Номер и дата исполнительного документа</p> <p>Наименование должника (ОПФ и Наименование организации/ФИО физического лица)</p> <p>ИНН и КПП должника</p> <p>ОГРН должника</p> <p>СНИЛС должника</p>

Вид документа	Извлекаемые факты
По умолчанию (рекомендуется использовать для документов, которые не удалось классифицировать)	Дата рождения должника Номер р/с должника для взыскания задолженности ФИО судебного пристава-исполнителя Должность судебного пристава-исполнителя Главное управление структурного подразделения судебного пристава-исполнителя Структурное подразделение судебного пристава-исполнителя Телефон структурного подразделения судебного пристава-исполнителя Сумма взыскания Валюта взыскания
	Дата документа ОПФ и Наименование организации/ФИО индивидуального предпринимателя ИНН и КПП организации/индивидуального предпринимателя ОГРН юридического лица ОГРНИП физического лица Сумма документа Валюта документа ФИО персоны Номер телефона Адрес электронной почты

Ограничения при извлечении

Ознакомьтесь с требованиями и ограничениями, которые нужно учесть при извлечении [печатей и подписей](#) и [фактов из скан-копий документов](#).

> Печатей и подписей

При извлечении из документов печатей и подписей необходимо учесть, что:

- обрабатываются только круглые печати размером от 30 до 60 миллиметров;
- обрабатываются только синие печати. Подписи в документе могут быть любого цвета;
- нельзя определить, являются ли извлеченные печати и подписи разными или одинаковыми;
- нельзя определить, принадлежат ли печати и подписи организациям и лицам, указанным в документе.

➤ Фактов из скан-копии документа

Чтобы *факты* корректно извлекались из скан-копий бумажных документов, убедитесь, что они соответствуют требованиям:

Требование	Возможные проблемы, если требование не соблюдено
Рекомендуемое разрешение изображения – 300-400 DPI	Не распознаны отдельные слова или символы или распознаны некорректно
Качество печати оригинального документа обеспечивает читабельность символов	
Текст документа имеет нулевой или положительный кернинг (межбуквенный интервал)	
Текст документа имеет как минимум единичный интерлиньяж (междустрочный интервал)	
Текст документа не перекрыт графическими элементами: подписями, печатями, штампами, резолюциями и т.п.	
Скан-копия документа не содержит физических повреждений: загибов, пятен, отверстий от дыроколов, степлерных скоб, скрепок и т.п.	Не распознаны отдельные слова или символы или распознаны некорректно. При загибе угла страницы возможно некорректное определение формата таблиц документов, что приводит к некорректному извлечению фактов из таблиц
Таблицы имеют замкнутую границу. Границы таблиц и ячеек хорошо пропечатаны и не имеют пропусков	Таблица не извлечена в структурированном виде. Пропущены некоторые ячейки таблицы
Ячейки таблицы не содержат графических элементов: подписей, печатей, штампов, резолюций и т.п.	Содержание ячейки не распознано или распознано некорректно
Паспорт на странице расположен параллельно границам изображения	Не извлечены факты из паспорта или извлечены некорректно
Изображение документа имеет сбалансированную яркость и контрастность в части четкости и цветности изображения	Не распознаны отдельные слова или символы или распознаны некорректно Не извлечены факты из паспорта или извлечены некорректно
Скан-копия содержит изображение только одного документа жесткой формы. Например, только паспорта	Извлечены факты только из одного документа Другой документ удален

Список распознаваемых языков

Сервисы Directum Ario поддерживают распознавание одноязычных и мультязычных документов.

По умолчанию включено распознавание документов на русском и английском языках. Чтобы добавить другой язык, используйте параметр **Languages**. Подробнее см. раздел «Как добавить язык распознавания».

Язык	Код языка	В стандартной поставке
Азербайджанский	aze	
Албанский	sqi	
Английский	eng	Да
Армянский	hye	
Белорусский	bel	Да
Болгарский	bul	
Венгерский	hun	
Грузинский	kat	
Датский	dan	
Иврит	heb	
Индонезийский	ind	
Испанский	spa	
Итальянский	ita	
Казахский	kaz	Да
Киргизский	kir	
Китайский	chi_sim	
Корейский	kor	
Латышский	lav	
Литовский	lit	
Монгольский	mon	
Немецкий	deu	
Норвежский	nor	
Польский	pol	
Португальский	por	
Румынский; молдавский	ron	
Русский	rus	Да
Сербский	srp	
Словацкий	slk	
Таджикский	tgk	

Язык	Код языка	В стандартной поставке
Тайский	tha	
Татарский	tat	
Турецкий	tur	
Узбекский	uzb	
Украинский	ukr	Да
Финский	fin	
Французский	fra	Да
Хорватский	hrv	
Чешский	ces	
Шведский	swe	
Эстонский	est	
Японский	jpn	