



Система интеллектуального анализа документов

Руководство администратора

Всего листов: 62

Аннотация

Настоящий документ представляет собой руководство администратора (Руководство) по работе с Системой интеллектуального анализа документов (Система) и является частью эксплуатационной документации на Систему.

Документ включает в себя:

- общие сведения о функциональных возможностях Системы;
- описание порядка подготовки инфраструктуры к развертыванию программного обеспечения Системы;
- описание порядка развертывания программного обеспечения Системы;
- описание операций по администрированию Системы.

Содержание

Аннотация	2
1. Введение	5
1.1. Область применения системы	5
1.2. Краткое описание возможностей системы	5
1.3. Уровень подготовки пользователя	5
1.4. Перечень эксплуатационной документации для ознакомления	5
1.5. Глоссарий	6
2. Общее описание системы «Система интеллектуального анализа документов»	11
2.1. Назначение системы	11
2.2. Условия применения системы	11
2.2.1. Общие условия	11
2.2.2. Требования к инфраструктуре	11
2.2.3. Требования к рабочей станции для развертывания контура	12
2.2.4. Требования к программному обеспечению инфраструктуры	13
3. Подготовка к работе системы «Система интеллектуального анализа документов»	15
3.1. Установка серверной части	15
3.1.1. Подготовка виртуальных машин и базовой среды	15
3.1.2. Развертывание инфраструктуры сервисов	16
3.1.3. Развертывание компонентов в Kubernetes	17
3.1.4. Настройка серверов и развёртывание БЯМ	17
3.1.5. Настройка мониторинга и проверка работоспособности	17
3.2. Порядок запуска	18
3.3. Порядок остановки	18
3.4. Порядок установки обновлений системы	19
3.5. Порядок проверки работоспособности системы	19
3.6. Решения по диагностированию системы	19
4. Описание операций системы «Система интеллектуального анализа документов»	20
4.1. Управление учетными записями пользователей	20
4.1.1. Общие сведения	20
4.1.2. Создание учетной записи пользователя	21
4.1.3. Блокировка учетной записи пользователя	27
4.1.4. Удаление учетной записи пользователя	28
4.2. Настройка процесса пакетной загрузки файла	29
4.3. Запуск загрузки файлов вне расписания	33
4.4. Пакетная загрузка файлов	33
4.5. Восстановление удаленных документов	33
4.6. Первичная настройка конфигурационных параметров процесса интеллектуальной обработки документа	34
4.7. Настройка процесса интеллектуальной обработки документа	35
4.8. Работа со справочником обязательных разделов документа	57
5. Планово-предупредительные мероприятия	59
5.1. Документирование результатов	60
6. Аварийное восстановление системы «Система интеллектуального анализа документов»	61
6.1. Критерии аварийного события	61
6.2. Порядок действий при выявлении аварийного события	61
6.3. Общие рекомендации по восстановлению деятельности системы «Система интеллектуального анализа документов»	62

1. Введение

1.1. Область применения системы

Областью применения Системы является деятельность по подготовке и актуализации текстов документов.

1.2. Краткое описание возможностей системы

Система предоставляет следующие возможности:

- поиск по тексту документа ссылок на нормативно-правовые акты (НПА) и ГОСТ, а также последующее определение актуальности данных ссылок;
- определение возможности подписания документа электронной подписью;
- проверка названий документов (НПА), на которые есть ссылки по тексту документа;
- проверка наличия приложений к основной части документа, которые упоминаются по тексту;
- анализ структуры документа на предмет наличия обязательных разделов;
- при наличии в документе раздела «Список ссылок», определяется полнота такого списка в соответствии с составом ссылок, упоминаемых по тексту.

1.3. Уровень подготовки пользователя

Для выполнения операций, предусмотренных настоящим Руководством, пользователю с ролью «Администратор» потребуются знания основ информационной безопасности, навыки администрирования серверов и вычислительных сетей, рекомендуется наличие навыков и опыта работы с технологиями docker, kubernetes, Keycloak, Postgresql, Arangodb, а также опыт работы с системами управления документами.

Пользователю с ролью «Инженер ИИ» необходимо обладать опытом управления системами с искусственным интеллектом (ИИ) и опытом настройки Большой языковой модели (БЯМ), что предполагает наличие навыков создания, редактирования и оптимизации инструкций для БЯМ, навыков оценки качества выходных данных БЯМ, навыков анализа текстов для составления инструкций и примеров для БЯМ. Также потребуется понимание принципов обработки естественного языка, понимание синтаксиса, семантики и языковой структуры русского и английского языков.

1.4. Перечень эксплуатационной документации для ознакомления

Перед началом работы по установке, настройке, администрированию Системы пользователю с ролью «Администратор» необходимо ознакомиться с настоящим Руководством в полном объеме.

Для настройки проверок интеллектуальной верификации документов пользователю с ролью «Инженер ИИ» при изучении Руководства необходимо особое обратить внимание на п. 4.6 Руководства.

1.5. Глоссарий

В тексте настоящего документа представлены следующие сокращения (см. **Таблица 1-1**).

Таблица 1-1 — Перечень сокращений

Сокращение/аббревиатура	Значение
АРМ	Автоматизированное рабочее место пользователя Системы
БД	База данных - совокупность данных, организованных в соответствии с концептуальной структурой, описывающей характеристики этих данных и взаимоотношения между ними, которая поддерживает одну или более областей применения
БД Реестра документов	База данных Реестра документов - Реляционная БД, содержащая атрибуты, характеризующие статус, в котором находится тот или иной документ в Системе. Запись данных в БД производится по результатам различных операций обработки документа. Чтение информации из этой БД используется для определения доступных сценариев обработки документа на различных стадиях его жизненного цикла в Системе
БЯМ	Большая языковая модель - модель машинного обучения, которая способна понимать и генерировать текст на естественном языке
ГОСТ	Государственный стандарт
ИИ	Искусственный интеллект
ИС	Информационная система
ИОИ	Интеллектуальная обработка информации
НПА	Нормативный правовой акт
ОС	Операционная система
ПО	Программное обеспечение

Сокращение/аббревиатура	Значение
Руководство	Руководство по работе с Системой интеллектуального анализа документов (руководство администратора Системы) – настоящий документ
Система	Система интеллектуального анализа документов
СУБД	Система управления базами данных - Совокупность программных и лингвистических средств общего или специального назначения, обеспечивающих управление созданием и использованием баз данных (БД).
ТУЗ	Техническая учетная запись
ЦОД	Центр обработки данных
API	Application Programming Interface - Программный интерфейс, описание способов взаимодействия одной компьютерной программы с другими, используется при разработке приложений
CRC	Циклический избыточный код
FTP	File Transfer Protocol - Протокол, относящийся к прикладному уровню и отвечающий за передачу данных между двумя Системами
JSON	JavaScript Object Notation - Текстовый формат обмена данными, основанный на языке программирования JavaScript
LLM	Large Language Model, Большая языковая модель
TCP/IP	Transmission Control Protocol/Internet Protocol - Набор сетевых протоколов передачи данных, используемых в сетях, включая глобальную сеть Интернет
URL	Uniform Resource Locator - система унифицированных адресов электронных ресурсов, или единообразный определитель местонахождения ресурса
VM	Виртуальная машина

В тексте настоящего документа представлены следующие термины и определения (см. **Таблица 1-2**).

Таблица 1-2 — Перечень терминов и определений

Термин	Определение
Авторизация	Предоставление определенному лицу или группе лиц прав на выполнение определенных действий
Аутентификация	Проверка подлинности предъявленного пользователем идентификатора
Верификация (интеллектуальная верификация)	Проверка документа, в том числе с использованием технологий искусственного интеллекта, в соответствии с правилами, указанными в Техническом задании
Директория	Указатель пути к одному или нескольким файлам на FTP-сервере, то есть, адрес файла
Документ, электронный документ	В целях настоящего документа термин используется в следующих значениях: вне Системы – текстовый документ в формате *.pdf, *.doc, *.docx, *.rtf, *.txt, *.csv, *.odt, *.xls, *.xlsx, содержащий в том числе нормативные ссылки на другие документы. Носителем документа является файл (файл с документом); в Системе – сущность, объединяющая все элементы и структуры хранения, которые содержат информацию (в том числе обработанную с использованием искусственного интеллекта) из файла с документом
Заказчик	Физическое или юридическое лицо, которое инициирует, заказывает и финансирует разработку, модификацию или поддержку программного обеспечения Системы. После приемки услуг по разработке Системы Заказчик является Владелец Системы
Идентификатор пользователя	Уникальный признак объекта, позволяющий отличать его от других
Идентификация	Процедура, в результате выполнения которой для субъекта выявляется его идентификатор, однозначно определяющий этого субъекта в Системе
Компонент (программного обеспечения)	Составная часть программного обеспечения, выполняющая определенную функцию

Нод (Node) (kubernetes)	Ноды или узлы, виртуальные или физические машины, на которых разворачивают и запускают контейнеры. Совокупность нод образует кластер kubernetes
Объектное (файловое) хранилище	Технология для удобного и надежного хранения большого объема данных
Пакетная загрузка документов	Режим пакетной загрузки (параллельной или последовательной) однотипных документов из директории FTP-сервера во внутреннее файловое хранилище Системы
Под (Pod) (в kubernetes)	Группа контейнеров с общими разделами, которые запускаются как одно приложение
Пользователь	Работник, наделенный правами доступа к Системе интеллектуального анализа документов
Правило	Набор критериев, на соответствие которым проверяется документ или раздел документа с помощью запроса к Большой языковой модели
Проверка	Процесс анализа и оценки документа на соответствие ожиданиям и потребностям пользователей с помощью правил обращения к Большой языковой модели
Распознавание содержания электронных документов	Преобразование текста и изображений, содержащихся в загруженных в Систему файлах документов, в машиночитаемые текстовые данные
Распознанный документ	Состояние электронного документа, которое образуется по результатам успешного проведения процедуры распознавания содержания
Реестр документов	Сущность, являющаяся совокупностью всех документов Системы интеллектуального анализа документов (см. также термин «Документ»)
Ролевая модель	Распределение прав доступа и обязанностей между ролями. Определяет, какие действия может выполнять каждая роль в рамках работы с Системой, включая создание и редактирование правил, проведение проверок и управление
Роль	Набор полномочий, который необходим пользователю для выполнения определённых рабочих задач. Каждый сотрудник может

	иметь одну или несколько ролей, а каждая роль может содержать от одного до множества полномочий. Роли пользователей в Системе: «Пользователь»; «Администратор»; «Инженер ИИ»
Роль «Администратор»	Главная роль Системы, предоставляющая доступ к настройкам Системы
Роль «Инженер ИИ»	Роль для лица, осуществляющего настройку правил по верификации документов
Роль «Пользователь»	Роль для лица, использующего функциональность Системы
Файл	Цифровой носитель информации, содержащий текстовую и графическую информацию
Docsplit	API-сервис, принимающий запрос от обработчика реестра событий и возвращающий разбитый по чанкам текст
Docsplit_consumer	Компонент, читающий таблицу событий РКС и отправляющий запрос в API-сервис Docsplit
Helm	Пакетный менеджер для kubernetes, который используется для управления приложениями в кластере
KubectI	Командная строка для управления Kubernetes; необходим для взаимодействия с кластером Kubernetes
Kubernetes	Платформа для автоматизации развёртывания, масштабирования и управления контейнеризированными приложениями. Поддерживает основные технологии контейнеризации (docker, rocket) и аппаратную виртуализацию
PostgreSQL	Объектно-реляционная СУБД с открытым исходным кодом, которая использует и расширяет язык SQL в сочетании со многими функциями, которые позволяют безопасно хранить и масштабировать самые сложные рабочие нагрузки с данными
sshpas	Утилита для автоматизации ввода пароля при использовании SSH, что требуется для настройки доступа к серверам
System_check	API-сервис, принимающий запрос от обработчика реестра событий и возвращающий извлеченные атрибуты Для
System_check_consumer	Компонент, читающий таблицу событий РКС и отправляющий запрос в API-сервис System_check

2. Общее описание системы «Система интеллектуального анализа документов»

2.1. Назначение системы

Система предназначена для решения следующих взаимосвязанных задач:

- распознавание содержания документов следующих форматов: *.pdf, *.doc, *.docx, *.rtf, *.txt, *.csv, *.odt, *.xls, *.xlsx;
- обработка текста документа, в том числе с применением технологий ИИ;
- обеспечение систематизированного хранения документов и связей между ними.

Обобщенным назначением Системы является информационная поддержка пользователя в принятии решений о необходимости редактирования текста документа.

2.2. Условия применения системы

2.2.1. Общие условия

Режим эксплуатации Системы: 5x8 (5 дней в неделю по 8 часов) в часы работы Заказчика.

Для возможности администрирования Системы должны выполняться условия, предусмотренные пп. 2.2.2, 2.2.3, 2.2.4 Руководства.

Для настройки журналирования событий по информационной безопасности уровень логирования рекомендуется выставлять в warning - как компромиссный уровень между объемом информации и детализацией.

2.2.2. Требования к инфраструктуре

Для работы Системы требуется 10 виртуальных машин, состав которых определен в **Таблице 2-1** настоящего Руководства.

Таблица 2-1 – Состав виртуальных машин

Компонент	Состав	Название сервера	ip	Примечание
Центр анализа – Большая языковая модель	code LLM, nvidia GPU drivers, docker-toolkit docker engine	vdc-rk-llm-gpu	10.10.10.10	-
СУБД	PostgreSQL	vdc-rk-postgres01	10.10.10.11	-
Master node kubernetes 01	k8s	vdc-rk-k8s01	10.10.10.12	-

Worker node kubernetes 02	k8s	vdc-rk-k8s02	10.10.10.13	-
Worker node kubernetes 03	k8s	vdc-rk-k8s03	10.10.10.14	-
Worker node kubernetes 04	k8s	vdc-rk-k8s04	10.10.10.15	-
СУБД Графовой модели данных	ArangoDB	vdc-rk-arango	10.10.10.16	-
FTP-сервер загрузки документов	vsftpd	vdc-rk-ftp01	10.10.10.17	-
FTP-сервер загрузки правил	vsftpd	vdc-rk-ftp02	10.10.10.18	-
Сервер объектного хранилище S3	minio	vdc-rk-minio01	10.10.10.19	-

2.2.3. Требования к рабочей станции для развертывания контура

Для выполнения операций по развертыванию контура Системы на рабочей станции должны быть выполнены следующие требования:

- операционная система Linux:
 - рабочая станция должна быть запущена на базе операционной системы (ОС) Linux;
 - ОС должна взаимодействовать с драйверами на GPU;
 - дополнительные требования к ОС определяются внутренними документами Заказчика;
- программное обеспечение:
 - helm;
 - kubectl;
 - sshpass;
- сетевые настройки: рабочая станция должна иметь доступ по SSH ко всем серверам, участвующим в развертывании контура Системы;

- доступ по SSH-ключам: для упрощения процесса развертывания рекомендуется настроить беспарольный доступ по SSH с рабочей станции на все серверы;
- права доступа: пользователь на рабочей станции должен иметь привилегированный доступ (sudo) для выполнения административных задач;
- настройки безопасности: SSH-ключи, используемые для доступа к серверам, должны быть защищены и безопасно храниться;
- дополнительные утилиты: утилиты `wget`, `curl`, `tar`, `unzip`, `telnet nc net-tools` должны быть подготовлены для скачивания и распаковки файлов;
- резервное копирование: рекомендуется регулярно создавать резервные копии конфигурационных файлов и ключей на рабочей станции;
- проверка и тестирование: на рабочей станции должны быть установлены инструменты для тестирования и проверки конфигураций, такие как ping, telnet, nslookup, net-tools, nc;
- документация: рабочая станция должна иметь доступ к актуальной документации о Системе и инструкциям по развертыванию.

2.2.4. Требования к программному обеспечению инфраструктуры

Для обеспечения работы Системы необходимо соблюдать следующие требования:

- наличие программного обеспечения (ПО) на виртуальных машинах (VM), составляющих инфраструктуру Системы, соответствующего требованиям, установленным в п. 2.2.4 Руководства;
- ОС: поддерживает драйверы GPU Nvidia H100 и соответствует техническим требованиям Заказчика по RHEL;
- требования к виртуальным машинам – предоставляются отдельно;
- в контуре должен быть установлен NTP-сервер;
- все машины контура должны синхронизировать время;
- на всех хостах должен быть заведён служебный пользователь с правом выполнения команд с привилегиями суперпользователя для установки и настройки компонентов Системы;
- на всех хостах должен быть установлен и настроен ssh с возможностью авторизации служебным пользователем;
- CPU – рекомендуется Xeon Gold или аналог;
- отсутствие VPN соединений между VM;

- сеть:
 - скорость не менее 1 Гбит;
 - агрегация физических линков на серверах для отказоустойчивости (не менее 2 подключений);
 - для схемы размещения в одном ЦОД все серверы должны быть в одной подсети без фильтрации трафика между ними;
- на каждой виртуальной машине должен быть установлен и настроен iptables;
- наличие Nexus: в Nexus должны быть созданы по списку репозитории для apt, helm, npm, maven, docker, pypi;
- наличие docker для запуска контейнеризованных приложений, необходимый для работы с docker-образами и контейнерами.

3. Подготовка к работе системы «Система интеллектуального анализа документов»

3.1. Установка серверной части

3.1.1. Подготовка виртуальных машин и базовой среды

Необходимо выполнить следующие шаги:

- а) Создать виртуальные машины:
 - FTP-серверы: две виртуальные машины для развертывания FTP-серверов (vsftpd) для загрузки файлов пользователями;
 - PostgreSQL-сервер: одна машина для PostgreSQL для хранения данных;
 - kubernetes-кластер: три виртуальные машины для kubernetes (K8s), на которых будет развёрнут кластер для компонентов обработки данных;
 - S3-хранилище: один сервер для MinIO, который обеспечит функциональность S3-хранилища;
 - Сервер с GPU: одна виртуальная машина с NVIDIA GPU для запуска контейнеров с поддержкой графического ускорителя, необходимых для компонента БЯМ;
 - ArangoDB-сервер: отдельный сервер для СУБД ArangoDB, предназначенной для специфичных задач управления данными.
- б) Установить ОС и сетевые настройки:
 - 1) установить совместимую с Linux ОС на всех виртуальных машинах (например, Ubuntu 20.04 или CentOS 7/8);
 - 2) настроить сеть, чтобы обеспечить доступ между серверами, а также открыть порты для необходимых сервисов:
 - PostgreSQL: 5432;
 - FTP: 21, 20;
 - kubernetes API: 6443;
 - kubernetes NodePorts: 30000, 30001, 30888, 30083, 30005, 30051, 30050;
 - MinIO: 9000;
 - ArangoDB: 8529;
 - 3) установить базовые утилиты и зависимости, включая curl, wget, git, vim, docker, kubectl, и другие необходимые инструменты.

3.1.2. Развертывание инфраструктуры сервисов

Необходимо выполнить следующие шаги:

- a) FTP-серверы (vsftpd):
 - 1) установить vsftpd на двух виртуальных машинах;
 - 2) настроить пользователей и директории для загрузки файлов;
 - 3) настроить безопасность и разрешения для FTP-доступа;
 - 4) настроить права для двух пользователей admin и user, где пользователь admin является администратором и имеет доступ к обеим папкам, пользователь user имеет доступ только к одной своей папке на одном FTP-сервере для загрузки документов.
- b) PostgreSQL:
 - 1) установить PostgreSQL на выделенной машине;
 - 2) настроить конфигурацию сети (разрешить доступ через localhost и IP);
 - 3) создать необходимые базы данных (БД) и пользователей.
- c) Kubernetes-кластер:
 - 1) установить kubernetes на четырех виртуальных машинах (используя kubeadm);
 - 2) настроить кластер и проверить его работоспособность;
 - 3) установить Local Path Provisioner для динамического предоставления PV (Persistent Volumes) на узлах кластера.
- d) S3-хранилище (MinIO):
 - 1) установить и настроить MinIO на выделенном сервере;
 - 2) создать необходимые бакеты для хранения данных;
 - 3) настроить доступы для пользователей и сервисов.
- e) ArangoDB:
 - 1) установить ArangoDB на выделенной машине;
 - 2) настроить сеть и параметры безопасности (пароли, доступы);
 - 3) подготовить базы данных и коллекции для работы приложений.
- f) Сервер с GPU:
 - 1) установить NVIDIA драйверы, nvidia-container-toolkit и docker;
 - 2) настроить docker для работы с GPU;
 - 3) проверить работу GPU с помощью nvidia-smi и запуска тестового контейнера.

3.1.3. Развертывание компонентов в Kubernetes

Необходимо выполнить следующие шаги:

- a) Загрузить образы:
 - 1) подготовить образы компонентов интеллектуальной обработки информации (ИОИ), DWH, OCR, User Check в формате .tgz;
 - 2) загрузить подготовленные образы в Nexus после настройки Nexus как docker registry.
- b) Nexus:
 - 1) развернуть Nexus в kubernetes;
 - 2) настроить проху и hosted-репозитории;
 - 3) импортировать образы docker компонентов в Nexus.
- c) Установить компоненты – развернуть компоненты ИОИ, DWH, OCR, User Check в Kubernetes, используя манифесты YAML и предварительно загруженные образы из Nexus.
- d) Airflow, Prometheus, Metrics-server, Grafana:
 - 1) установить Airflow, Prometheus, Metrics-server и Grafana в kubernetes;
 - 2) импортировать DAG файлы в Airflow;
 - 3) проверить корректность работы всех компонентов, а также доступность компонентов через веб-интерфейсы.

3.1.4. Настройка серверов и развёртывание БЯМ

Необходимо выполнить следующие шаги:

- a) Настроить docker на GPU-сервере:
 - 1) установить и настроить docker для работы с nvidia-container-toolkit;
 - 2) проверить корректность настроек с помощью тестового GPU-контейнера.
- b) Запустить компоненты БЯМ:
 - 1) загрузить из Nexus образы компонентов БЯМ на сервер с GPU;
 - 2) запустить компоненты БЯМ в docker, используя настройки для работы с GPU.

3.1.5. Настройка мониторинга и проверка работоспособности

Необходимо выполнить следующие шаги:

- a) Prometheus Node Exporter и PostgreSQL Exporter:

- 1) установить Prometheus Node Exporter на всех серверах;
 - 2) установить PostgreSQL Exporter на сервере с базой данных PostgreSQL;
 - 3) настроить Prometheus для сбора метрик с node exporter и PostgreSQL exporter.
- b) Метрики:
- 1) проверить корректность поступления метрик всех сервисов в Prometheus;
 - 2) настроить визуализацию метрик в Grafana и создать необходимые дашборды.
- c) Выполнить финальную проверку:
- 1) проверить доступность всех сервисов, а также корректность работы сервисов;
 - 2) выполнить тестирование основных компонентов Системы:
 - загрузка и выгрузка файлов через FTP;
 - доступ и записи в базы данных PostgreSQL и ArangoDB;
 - корректная работа всех компонентов, развернутых в kubernetes;
 - 3) проверить доступность метрик в Prometheus и Grafana.

3.2. Порядок запуска

Запуск в работу любого компонента Системы необходимо выполнять в зависимости от типа установки, т.к. большинство компонентов находятся в kubernetes.

Запуск компонента, имеющего файл манифеста, необходимо выполнить следующим образом:

```
kubectl apply -f <manifest file yaml>
```

Если компонент разворачивается с помощью пакетного менеджера helm, то его запуск необходимо выполнить следующим образом:

```
helm install <имя helm релиза в kubernetes> <имя helm chart /  
репозитория> -f values.yaml --namespace <название kubernetes  
пространства имен, куда необходимо развернуть компонент>
```

3.3. Порядок остановки

Для остановки работы компонента необходимо исходить из типа установки/удаления, т.к. большинство компонентов находится в kubernetes.

Остановку компонента, имеющего файл манифеста, необходимо выполнить следующим образом:

```
kubectl delete -f < manifest file yaml>
```

Если компонент разворачивается с помощью пакетного менеджера helm, то его остановку/удаление необходимо выполнить следующим образом:

```
helm uninstall <имя helm релиза в kubernetes> --namespace  
<название kubernetes пространства имен, куда необходимо  
развернуть компонент>
```

3.4. Порядок установки обновлений системы

В случае предоставления новой версии кодовой базы компонентов Системы их обновление производится в соответствии с инструкциями по их установке – повторяется процедура развертывания компонентов. Предыдущие версии компонентов Системы будут заменены обновленными в связи с использованием kubernetes.

Компоненты Системы, развертываемые с помощью docker, также проходят процедуру в рамках инструкции по их развертыванию. Однако рекомендуется остановить компоненты с помощью следующей команды:

```
docker compose down --remove-orphans
```

3.5. Порядок проверки работоспособности системы

Способы проверки и диагностирования работы каждого компонента указаны в соответствующем пункте документа «Инструкция по развертыванию и установке программного обеспечения», в котором приведено описание развертывания компонента.

3.6. Решения по диагностированию системы

Способы проверки и диагностирования работы каждого компонента указаны в соответствующем пункте документа «Инструкция по развертыванию и установке программного обеспечения», в котором приведено описание развертывания компонента.

4. Описание операций системы «Система интеллектуального анализа документов»

4.1. Управление учетными записями пользователей

4.1.1. Общие сведения

Для управления учетными записями пользователей с ролью «Пользователь» необходимо последовательно выполнить следующие шаги:

- а) Войти в компонент управления доступом (Keycloak). Для этого в адресной строке браузера необходимо ввести соответствующий URL.
- б) На открывшейся странице (см. **Рис. 4-1**) необходимо ввести имя пользователя с ролью «Администратор» и пароль, нажать на кнопку входа «Sign In».

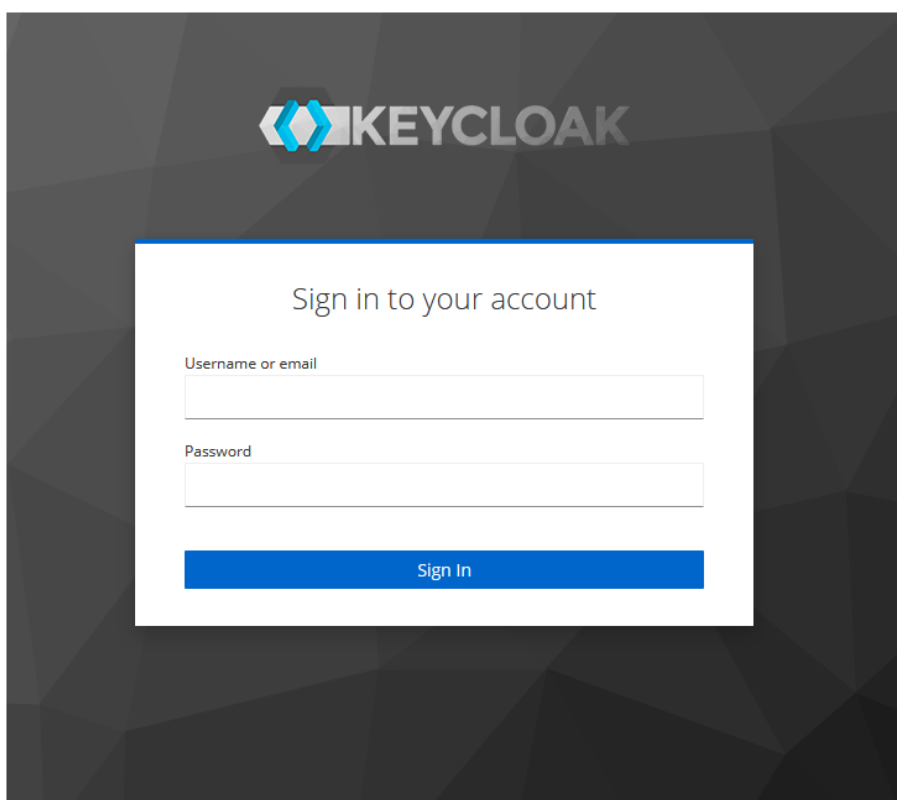


Рис. 4-1 – Страница авторизации в компоненте управления доступом к Системе

- с) В списке, расположенном в верхней части главного меню, выбрать пункт «Graph», который определяет зону действия прав пользователя в Системе (см. **Рис. 4-2**).

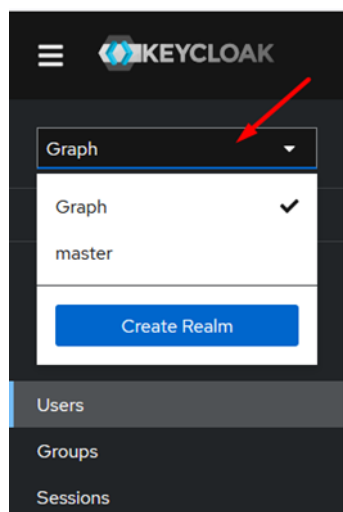


Рис. 4-2 – Выбор зоны действия прав пользователя

d) Выполнить любую из следующих доступных операций с учетными записями пользователей:

- 1) создание;
- 2) блокировка;
- 3) удаление.

4.1.2. Создание учетной записи пользователя

Для создания новой учетной записи авторизованному пользователю с ролью «Администратор» необходимо выполнить следующие действия.

- a) В разделе «Manage» главного меню выбрать пункт «Users».
- b) На открывшейся странице нажать на кнопку «Add user» (см. **Рис. 4-3**) для вызова анкеты для создания нового пользователя.

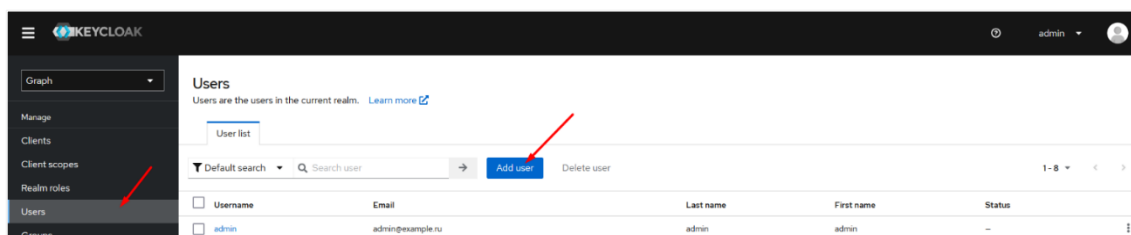


Рис. 4-3 – Вызов анкеты для создания нового пользователя

Ожидаемый результат – откроется страница анкеты для создания нового пользователя (см. **Рис. 4-4**).

Рис. 4-4 – Анкета для создания нового пользователя

- с) На открывшейся странице (см. **Рис. 4-4**) необходимо заполнить поля анкеты нового пользователя согласно данным, приведенным в **Таблице 4-1** настоящего Руководства.

Таблица 4-1 – Информация о заполнении анкеты для создания нового пользователя

Наименование поля	Обязательность	Рекомендации по заполнению
Required user actions	Нет	Не рекомендуется заполнять
Username	Да	Обязательно заполнить. Используется в качестве имени пользователя при авторизации в Системе
Email	Нет	Рекомендуется к заполнению. Может использоваться для восстановления или изменения пароля пользователем самостоятельно
Email verified	Нет	Допускается использование значения по умолчанию (отключено). Зависит от политики управления доступом, действующей в организации

First name	Нет	Рекомендуются к заполнению поля фамилии и имени сотрудника, который будет являться пользователем Системы. Отображается в профиле авторизованного пользователя
Last name	Нет	Рекомендуются к заполнению поля фамилии и имени сотрудника, который будет являться пользователем Системы. Отображается в профиле авторизованного пользователя

d) Нажать на кнопку «Join Groups».

Ожидаемый результат – открывается модальное окно для определения группы пользователя (см. **Рис. 4-5**).

e) В открывшемся модальном окне (см. **Рис. 4-5**) чек-боксом выделить запись «GraphUsers», которая определяет состав прав пользователей группы, и нажать на кнопку «Join».

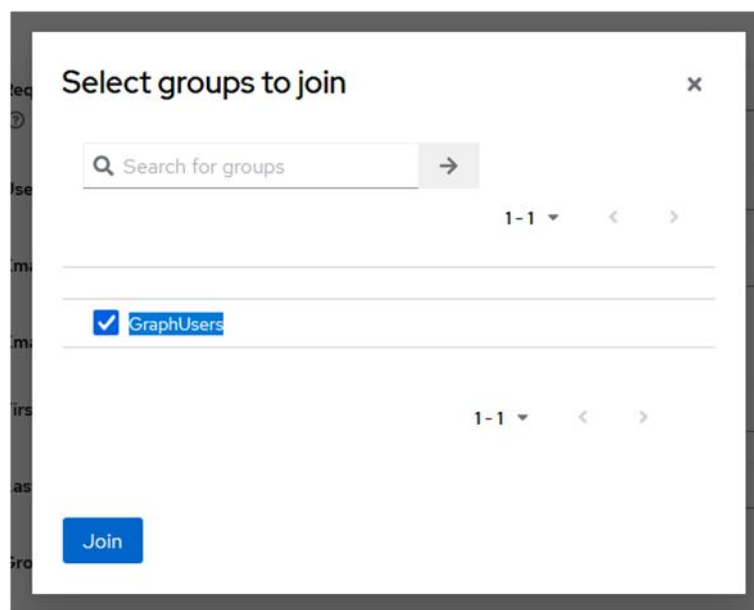


Рис. 4-5 – Модальное окно для определения группы пользователя

f) Нажать на кнопку «Create».

Ожидаемый результат – при полном заполнении анкеты отображается сообщение об успешном создании пользователя (см. **Рис. 4-6**) и открывается карточка нового пользователя.

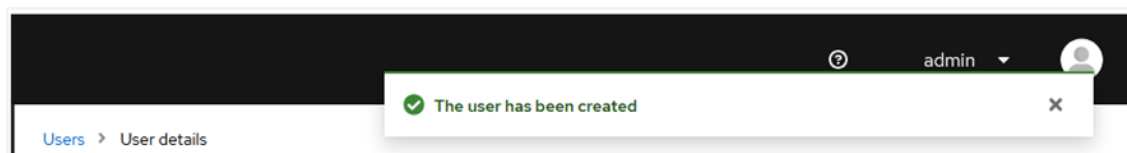


Рис. 4-6 – Сообщение об успешном создании пользователя

- g) В карточке пользователя открыть вкладку «Credentials».
- h) На открывшейся странице нажать на кнопку «Set password».

Ожидаемый результат – открывается модальное окно (см. **Рис. 4-7**).

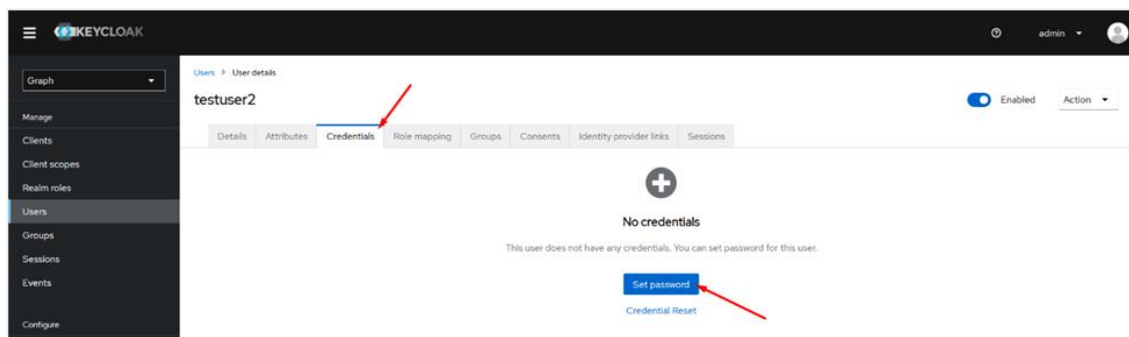


Рис 4-7 – Вызов функции создания пароля для нового пользователя

- i) В открывшемся модальном окне (см. **Рис. 4-8**) выполнить следующие действия:
 - 1) задать пароль для нового пользователя (поле «Password»);
 - 2) повторить ввод заданного пароля в поле «Password confirmation»;
 - 3) перевести переключатель управления режимом временного пароля («Temporary») в выключенное состояние (Off);
 - 4) нажать кнопку сохранения пароля «Save»;
 - 5) подтвердить необходимость задания пароля новому пользователю в открывшемся новом модальном окне.

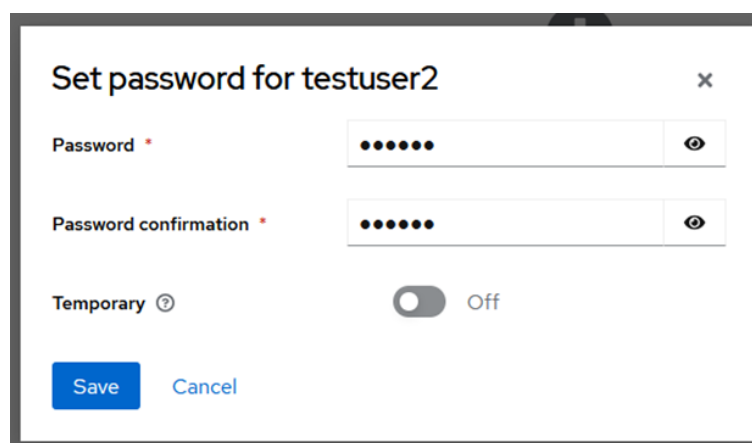


Рис. 4-8 – Модальное окно, подтверждение необходимости задания пароля новому пользователю

Ожидаемый результат – по факту завершения задания пароля новому пользователю будет выведено сообщение и на вкладке «Credentials» отобразится запись о создании пароля (см. **Рис. 4-9**).

*Примечание – в строке с записью о создании пароля содержится кнопка «Reset password». При необходимости смены пароля при нажатии на указанную кнопку вызывается модальное окно, представленное на **Рис. 4-8**.*

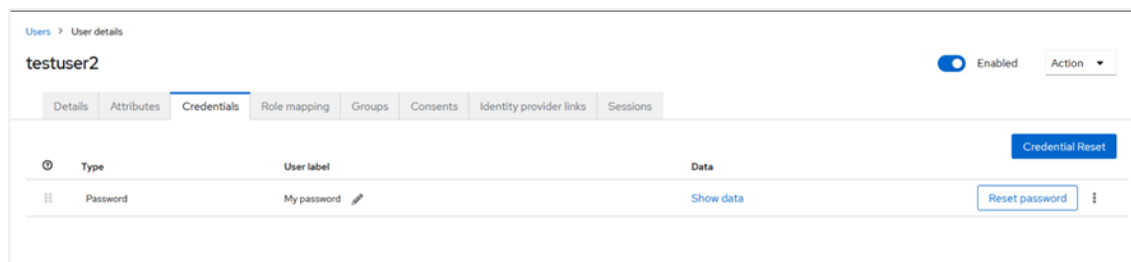


Рис. 4-9 – вкладка «Credentials», запись о создании пароля

После выполнения действий, перечисленных в пунктах «а)» – «і)» п. 4.1.2 Руководства будет создана учетная запись для нового пользователя Системы; информация о таком пользователе отобразится на странице списка пользователей (см. **Рис. 4-3**).

Примечание – для возможности идентификации действий пользователей по Журналу событий Системы каждому пользователю должна создаваться уникальная учетная запись. Заведение групповых или анонимных учетных записей не допускается.

Владелец Системы применяет организационно-распорядительные меры для регулирования процесса выдачи учетных записей.

Важно! Все пользователи группы «GraphUsers» (описание приведено в п. «е)» п. 4.1.2 Руководства) наследуют состав прав, необходимых и достаточных для работы в Системе. Полный перечень таких прав представлен в **Таблице 4-2**.

Таблица 4-2 – Состав прав пользователей с ролью «Пользователь»

№ п/п	Наименование полномочия	Описание полномочия
1	manageAccessLevel	Получение уровней доступа
2	basketAccess	Получение списка вершин из корзины
3	detailGraphAccess	Получение справки
4	documentAccess	Получение информации о документе
5	documentVersionAccess	Получение информации о версии документа
6	filterGraphAccess	Получение из метадаты списка доступных рёбер
7	refreshGraphAccess	Обновление графа

8	graphAccess	Запрос получения проекта (параметров формирования графа) по идентификатору
9	graphDepthAccess	Сохранение/получение глубины и направления графа
10	expandedVertexAccess	Раскрытие вершины графа
11	linkDocumentAccess	Получение информации о ссылках на упоминаемые документы
12	verificationAccess	Получение информации о проверках (список проверок и результаты проверок)
13	createProject	Создание проекта
14	projectUpdate	Обновление проекта
15	getGraphState	Получение графа из кэша по идентификатору
16	createGraphState	Создание графа в кэше
17	projectHistory	Получение текущего состояния проекта по идентификатору проекта
18	searchControllerGet	Получение списка недавно выполненных поисковых запросов
19	schemaAccess	Обновление схемы
20	manageAccessCreate	Создание коллекции через Админку
21	manageAccessGetAll	Получение списка коллекций через Админку
22	manageAccessGet	Получение коллекции по идентификатору через Админку
23	manageAccessUpdate	Обновление коллекции через Админку
24	manageAccessDelete	Удаление коллекции по идентификатору через Админку
25	searchAccess	Получение результата поискового запроса
26	manageAnimateGet	Получение анимаций
27	manageAnimate	Обновление анимации
28	manageFigureGet	Получение фигур
29	manageFigure	Обновление фигуры
30	manageFontGet	Получение шрифтов

31	manageFont	Обновление шрифта
32	manageIconGet	Получение svg иконки
33	manageIcon	Создание svg иконки
34	manageLineGet	Получение линий
35	manageLine	Создание линии
36	manageTextGet	Получение текста
37	manageText	Обновление текста
38	searchAccessEx	Получение поискового результата с использованием расширенного фильтра по атрибутам
39	getAllConfigurationAccess	Получение списка конфигураций
40	getAllConfigurationPaginationAccess	Получение списка конфигураций с пагинацией
41	getByIdConfigurationAccess	Получение конфигурации по id
42	saveConfigurationAccess	Сохранение конфигурации
43	updateConfigurationAccess	Обновление конфигурации
44	deleteConfigurationAccess	Удаление конфигурации
45	eventRegisterAccess	Создание событий обработки документа
46	s3UploaderAccess	Загрузка файлов через интерфейс пользователя

4.1.3. Блокировка учетной записи пользователя

Для блокировки учетной записи необходимо последовательно выполнить следующие шаги.

- а) В списке пользователей (пункт «Users» главного меню) выбрать необходимую запись и нажать на соответствующую строку списка.
- б) В открывшейся карточке пользователя в правом верхнем углу перевести переключатель «Enabled» (отмечен стрелкой на **Рис. 4-10**) в выключенное состояние.

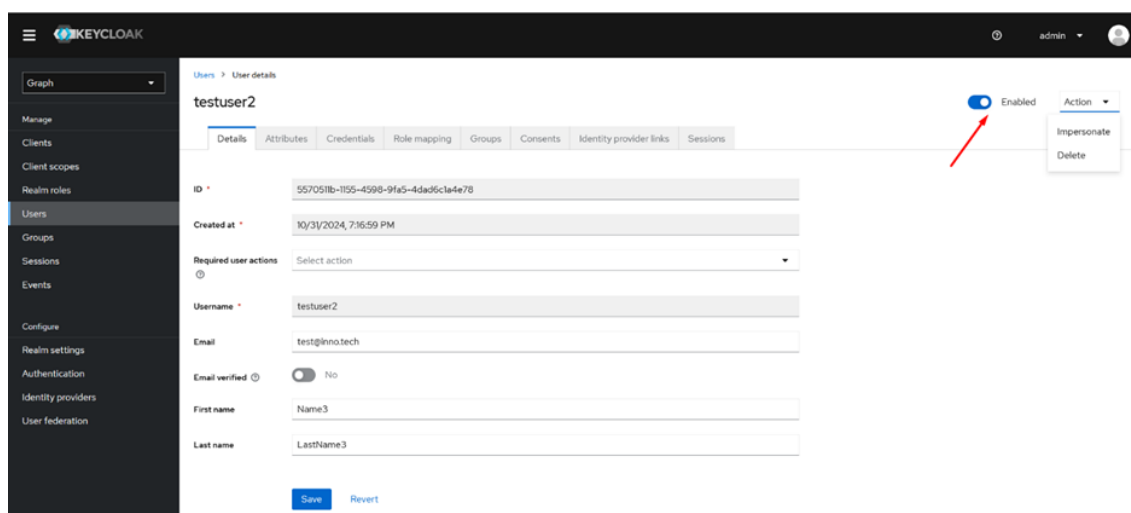


Рис. 4.10 – Блокировка учетной записи

- с) Дождаться появления сообщения о подтверждении сохранения выполненных действий (см. Рис. 4-11).

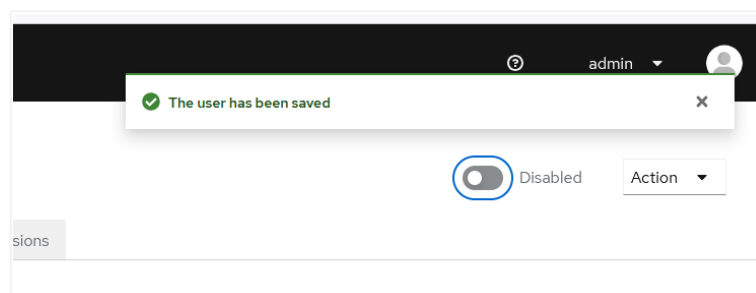


Рис. 4-11 – Переключатель блокировки учетной записи

По результатам выполненных действий, указанных в пп. «а)а)» – «с)» п. 4.1.3 Руководства, соответствующая учетная запись будет деактивирована (заблокирована), информация отобразится в поле «Status» списка пользователей (см. Рис. 4-12).

Users					
Users are the users in the current realm. Learn more					
User list					
Default search Search user Add user Delete user 11 - 13					
☐	Username	Email	Last name	First name	Status
☐	test456	test456@example.com	User	Test	–
☐	testuser	user@innotech.com	Name2	Name1	–
☐	testuser2	test@innotech.com	LastName3	Name3	Disabled

Рис. 4-12 – Отображение статуса заблокированной учетной записи

4.1.4. Удаление учетной записи пользователя

Для удаления учетной записи необходимо последовательно выполнить следующие шаги:

- а) В списке пользователей (пункт «Users» главного меню) выбрать необходимую запись.

б) В правой части строки вызвать контекстное меню (см. **Рис. 4-13**).

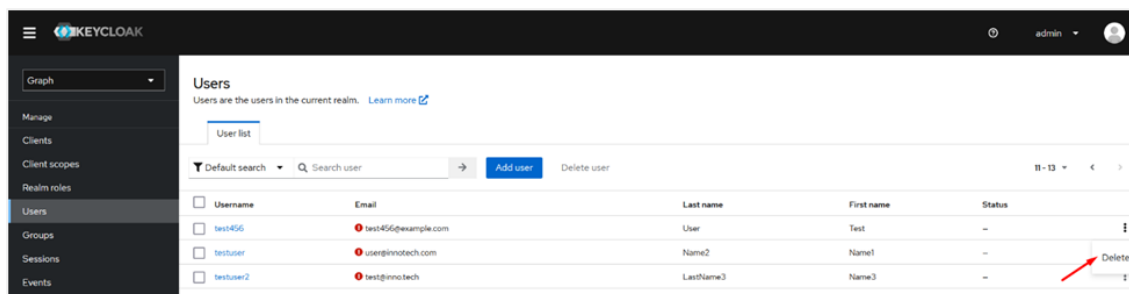


Рис. 4-13 – Удаление учетной записи

с) В открывшемся меню выбрать единственный пункт «Delete».

По результатам выполненных действий, указанных в пп. «а)» – «с)» п. 4.1.4 Руководства, соответствующая учетная запись будет удалена, отобразится сообщение об удалении учетной записи (см. **Рис. 4-14**). Также запись о пользователе не будет отображаться в списке пользователей.

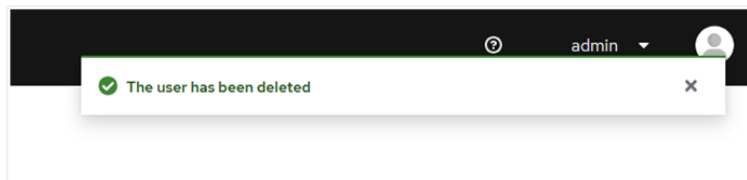


Рис. 4-14 – Сообщение об удалении учетной записи

Примечание – учетные записи для пользователей с ролью «Администратор» и «Инженер ИИ» создает суперпользователь вне Системы (описание приведено в пп. «4)» п. «а)» п. 3.1.2 Руководства).

4.2. Настройка процесса пакетной загрузки файла

После разворачивания Системы конфигурационные параметры пакетной загрузки файлов соответствуют значениям по умолчанию. Для их изменения можно использовать веб-интерфейс платформы kubernetes, набрав соответствующий URL в командной строке браузера.

После ввода URL платформы kubernetes необходимо выполнить следующие шаги:

а) На открывшейся вкладке браузера в левой части верхней строки (слева от окна поиска) в выпадающем окне выбрать необходимое пространство имен (rks-docs) (см. **Рис. 4-15**).

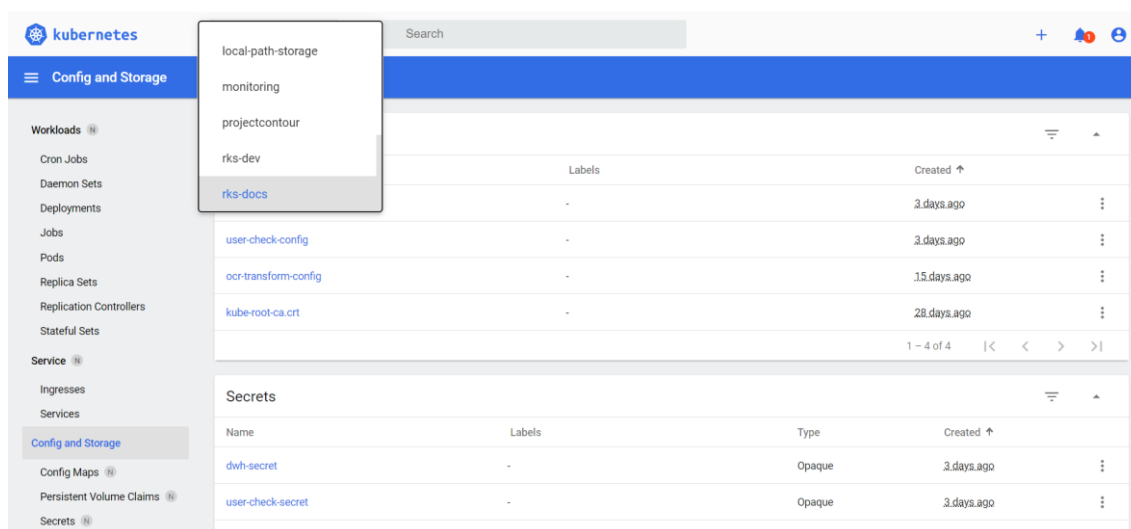


Рис. 4-15 – Выбор пространства имен

- б) Для просмотра текущих значений параметров конфигурации выбрать конфигурационную карту с именем dwh-config в разделе Config Maps (см. **Рис. 4-16**).

Примечание – для просмотра секретов выбрать список dwh-secrets в разделе Secrets (описание приведено в п. «д») п. 4.2 Руководства).

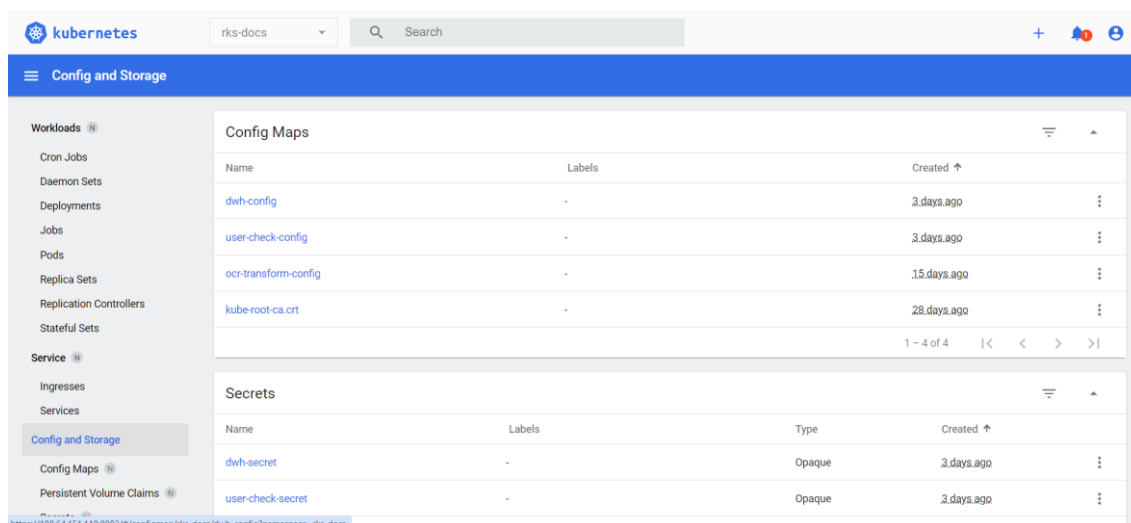


Рис. 4-16 – Выбор конфигурационной карты

В разделе Data на открывшейся странице (см. **Рис. 4-17**) доступны для просмотра значения следующих конфигурируемых параметров:

- «APPLICATION_NAME» – системный параметр, используемый для идентификации событий в едином реестре событий Системы;
- «FTP_HOST» – URL сервера FTP, с которого будут читаться документы в пакетном режиме;
- «FTP_PATH» – путь директории, в которой должны размещаться файлы для

пакетной загрузки в Систему (интерфейсной директории) на сервере FTP,

- «LOCAL_TEMP_PATH» – системный параметр, указывающий на локальную временную папку, используемую для переноса файлов в файловое хранилище Системы;
- «POSTGRES_DB» – имя БД Реестра документов Системы;
- «POSTGRES_HOST» – URL СУБД, поддерживающей функционирование Реестра документов Системы;
- «POSTGRES_PORT» – порт для подключения к СУБД, поддерживающей функционирование Реестра документов Системы;
- «S3_BUCKET» – имя папки в файловом хранилище Системы для пакетной загрузки в неё файлов с сервера FTP;
- «S3_HOST» – URL файлового хранилища Системы;
- «S3_PORT» – порт для подключения к файловому хранилищу Системы.

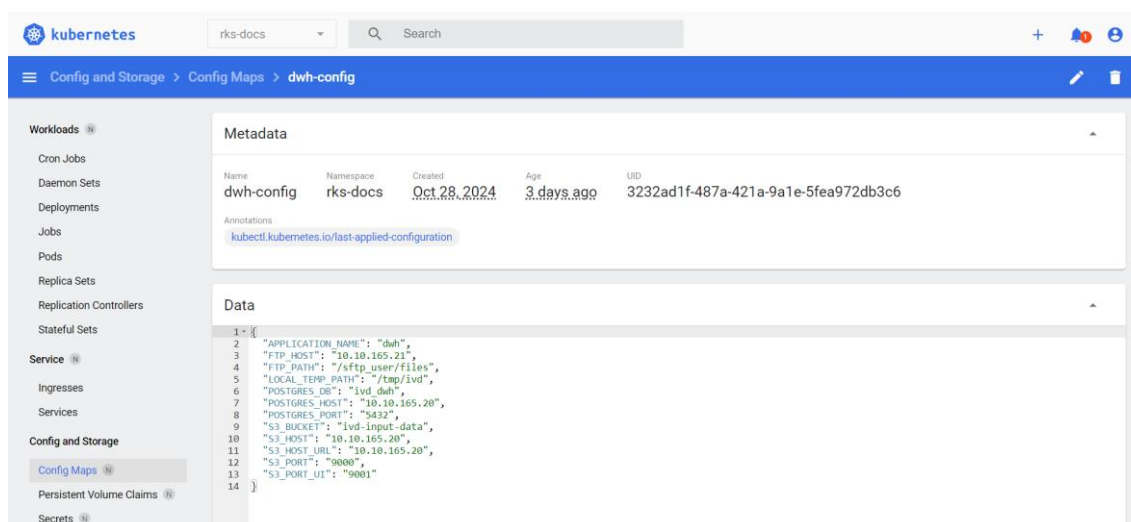


Рис. 4-17 – Просмотр конфигурационных параметров

с) Для редактирования значений, указанных конфигурационных параметров, выполнить следующие действия (см. **Рис. 4-18**):

- 1) нажать на кнопку с обозначением карандаша в правой верхней части справа вверху страницы;
- 2) в открывшемся всплывающем окне выбрать удобную форму представления конфигурационного файла;
- 3) внести необходимые изменения в раздел annotations и data;
- 4) сохранить изменения нажатием кнопки Update в левой нижней части всплывающего окна.

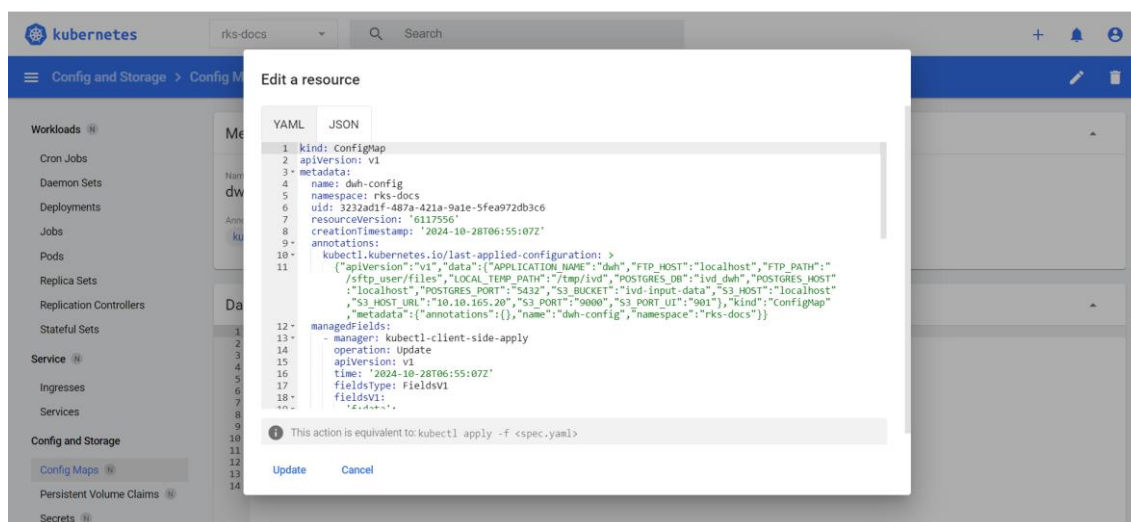


Рис. 4-18 – Редактирование конфигурационных параметров

д) Для редактирования паролей подключения к ресурсам, требуемым при пакетной загрузке, в меню левой части страницы выбрать раздел Secrets (см. **Рис. 4-19**).

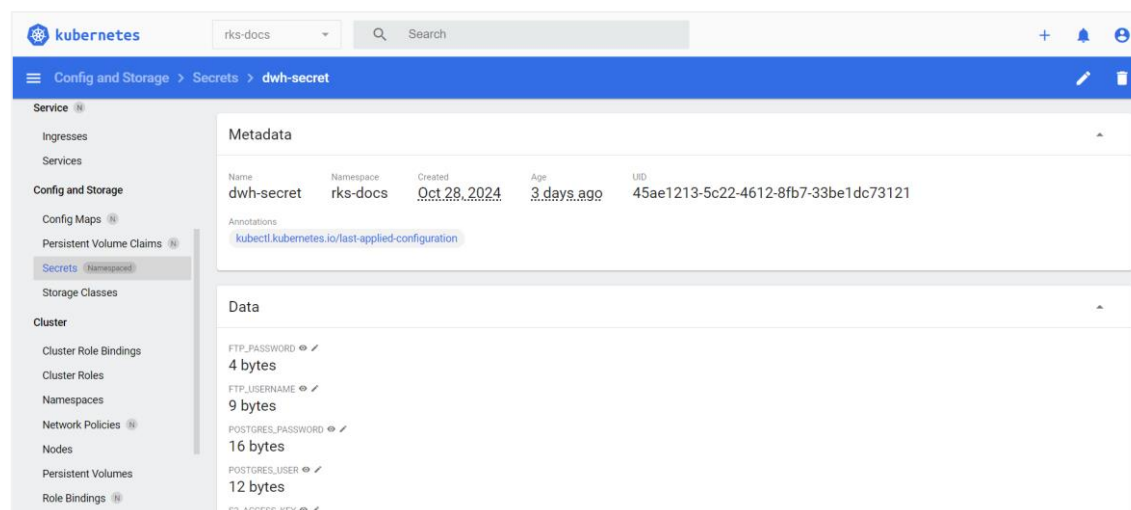


Рис. 4-19 – Раздел «Secrets»

В открывшейся вкладке, нажимая на кнопки просмотра и редактирования, размещенные в разделе Data правее названия параметра, можно просмотреть или изменить значения следующих секретов:

- «FTP_PASSWORD» – пароль подключения к серверу FTP;
- «FTP_USERNAME» – имя пользователя для подключения к серверу FTP;
- «POSTGRES_PASSWORD» – пароль для подключения к СУБД, поддерживающей функционирование Реестра документов Системы;
- «POSTGRES_USER» – имя пользователя для подключения к СУБД, поддерживающей функционирование Реестра документов Системы;
- «S3_ACCESS_KEY» – публичный ключ подключения к файловому хранилищу

Системы;

- «S3_SECRET_KEY» – пароль подключения к файловому хранилищу Системы.

4.3. Запуск загрузки файлов вне расписания

Для пакетной загрузки файлов вне расписания необходимо открыть оркестратор процессов обработки данных Системы, набрав соответствующий URL в командной строке браузера.

Открывшаяся страница веб-интерфейса оркестратора будет содержать один процесс с названием `load_pack`, который следует выбрать, и нажать на кнопку со стрелкой вправо в правом верхнем углу страницы. Таким образом будет запущен процесс пакетной загрузки данных из директории сервера FTP в файловое хранилище Системы, а также все последующие процедуры обработки документов (см. **Рис. 4-20**).



Рис. 4-20 – Запуск пакетной загрузки файлов

4.4. Пакетная загрузка файлов

Пакетная загрузка файлов в циклическом режиме опрашивает содержимое интерфейсной папки сервера FTP с интервалом 15 секунд без участия пользователя с ролями «Администратор» или «Пользователь». Таким образом, для проведения пакетной загрузки файлов достаточно загрузить их в интерфейсную директорию на сервере FTP с использованием любого из инструментов сетевого копирования файлов.

4.5. Восстановление удаленных документов

Для восстановления документа, который пользователь с ролью «Пользователь» случайно удалил из Системы (помечен признаком «Удален»), пользователю с ролью «Администратор» необходимо выполнить следующие действия:

- Войти в Систему как авторизованный пользователь с ролью «Пользователь»;
- В разделе «Файлы» с помощью инструмента поиска найти запись о файле, который содержит удаленный документ;
- В выбранной строке списка файлов вызвать контекстное меню, нажав на элемент «три точки» в крайней правой части строки;
- В меню выбрать первую позицию – «Перейти к документу»;

- е) Нажать на позицию «Перейти к документу» – будет осуществлен переход в карточку документа;
- ф) В адресной строке браузера скопировать идентификатор документа (doc_id), определяемый как последовательность символов, которая начинается после подстроки «document/V_DOC%2F» и заканчивается перед символом «/»;

Пример:

Адрес страницы: http://10.10.200.21:9090/document/V_DOC%2F06cba863-c8e6-44ce-a6c3-bacd6e9a8639/common

doc_id= '06cba863-c8e6-44ce-a6c3-bacd6e9a8639'

- г) Выполнить подключение к БД Системы;
- h) Направить в БД следующий запрос для восстановления файла:

```
update document.doc_registry
set
    deleted_dttm=null,
    deleted_by_user_id=null,
    actual_end_dt=null
where
    doc_id::text='идентификатор_документа_из_адресной_строки'
and
    deleted_dttm is not null
;
```

Необходимый документ появится в списке документов Системы.

4.6. Первичная настройка конфигурационных параметров процесса интеллектуальной обработки документа

После разворачивания Системы конфигурационные параметры сервиса интеллектуального определения атрибутов документа, реализуемого с использованием функциональности БЯМ, размещаются в файлах system_check.yaml и docsplit.yaml, расположенных на сервере FTP, по пути, определяемому в .env файле.

Конфигурационные параметры, используемые по умолчанию, определены в примерах конфигурационных файлов, расположенных в составе Системы по пути src/config/llm/system_check.yaml и src/config/llm/docsplit.yaml соответственно.

Секреты (ключи) и идентификаторы для доступа к FTP-серверу, базе данных Системы, интерфейсу БЯМ необходимо задать в .env файле, расположенном в каталоге с конфигурационным файлом docker-compose.yml, перед разворачиванием контейнера.

Требуемые секреты (ключи) и идентификаторы:

- «FTP_USERNAME» – имя пользователя для доступа к FTP»;
- «FTP_PASSWORD» – пароль пользователя для доступа к FTP;

- «EXT_DB_USERNAME» – имя пользователя для доступа к внешней БД Системы;
- «EXT_DB_PASSWORD» – пароль пользователя для доступа к внешней БД Системы;
- «PATH_CONFIG_SYSTEMCHECK» – путь до конфигурационного файла system_check;
- «PATH_CONFIG_DOCSPLIT» – путь до конфигурационного файла docsplit;
- «PATH_LOG_DIR» – путь до каталога хранения лог файлов компонента;
- «LLM_API_KEY» – ключ доступа к БЯМ (описание приведено в п. **Error! Reference source not found.** настоящего Руководства);
- «PATH_TO_MODEL» – абсолютный путь до файлов БЯМ.

Для изменения конфигурационных параметров сервиса БЯМ, включая URL серверов БЯМ и БД Реестра документов, следует загрузить новые (скорректированные) конфигурационные файлы в директорию /config/llm/ на сервере FTP любым из инструментов сетевого копирования файлов. Новые конфигурации будут применены при очередном обращении к сервису БЯМ.

При необходимости корректировки секретов (ключей) и идентификаторов доступа к серверу FTP, серверу БЯМ и серверу БД Реестра документов Системы изменения вносятся в указанные файлы и сервис разворачивается заново.

При внесении изменений в файлы конфигураций следует строго соблюдать рекомендации и условия назначения параметров, изложенные в комментариях перед соответствующим параметром. Часть параметров не должна изменяться в процессе работы сервиса. Все неизменяемые параметры отмечены в комментариях внутри файла.

4.7. Настройка процесса интеллектуальной обработки документа

Функциональность БЯМ по проведению интеллектуальной обработки документов настраивается в разделе attributes.

Для описания профиля используются общие и специальные (подраздел config) атрибуты. Назначение и наименования общих атрибутов профиля для всех профилей совпадают. Специальные атрибуты имеют различные наименования и функциональный смысл для каждой процедуры интеллектуальной обработки документа, что отражает её особенности. Изменяемые специальные атрибуты, кроме инструкции для модели (prompts), могут отсутствовать. И наоборот, при наличии специальных атрибутов при описании профиля, инструкции для модели могут отсутствовать.

В предоставленных конфигурационных файлах могут быть указаны параметры, не включенные в **Таблицы 4-3 – 4-9** настоящего Руководства. Такие параметры не допускается удалять и редактировать. Пояснения к таким параметрам приведены в самих конфигурационных файлах.

Списки изменяемых атрибутов приведены ниже в **Таблицах 4-3 – 4-10** настоящего Руководства. Изменяя состав списков можно настраивать контекст работы процедуры (например, для отнесения различных резюме к типу «Трудовые книжки» достаточно внести строку «Резюме» в параметр `work_records_doc_flg` – список ключевых образцов текста для отнесения документа к типу «Трудовая книжка»).

Общие правила обучения Системы для настройки извлечения информации из документов, требуемой для интеллектуальной обработки документов, указаны в **Таблицах 4-5 – 4-9** настоящего Руководства.

Таблица 4-3 – *Изменяемые атрибуты общего описания интеллектуальной обработки документов*

Параметр	Описание
<code>exception_doc_ids</code>	Список uuid документов (не версий), для которых проверка не проводится
<code>max_retry_count</code>	Максимальное количество попыток извлечения атрибута

Таблица 4-4 – *Изменяемые атрибуты общего описания профиля проверки (подраздел profiles)*

Параметр	Описание
Номер профиля	Обозначается натуральным числом
<code>is_default</code>	Флаг профиля по умолчанию
<code>updated_date</code>	Дата изменения/создания профиля

Таблица 4-5 – Общие правила обучения Системы и изменяемые специальные атрибуты для выделения ссылок на НПА и ГОСТ, которые содержатся в тексте документа (подраздел config)

Параметр	Разделы	Описание
Общие правила обучения Системы	-	Необходимо выполнить следующие шаги: – указать требуемые типы ссылок в параметре doc_types; – отредактировать указанный параметр в разделе common_params конфигурационного файла (см. Таблица 4-9); – указать необходимые исключения для типов ссылок в параметре not_links; – при получении нерелевантного результата при использовании новых типов специфичных документов – отредактировать подаваемые примеры в Систему в параметрах expected_outputs_first, expected_outputs_second, additional_tpn_examples, а также в инструкции в параметре prompt_templates: nra_extract_system_template:examples_instructions, следуя правилам, указанным для соответствующего параметра в настоящей таблице; – при получении нерелевантного результата при использовании новых типов специфичных документов – отредактировать параметры фильтрации результатов в разделе filters, следуя правилам, указанным для соответствующего параметра в настоящей таблице;

		– при получении нерелевантного результата при использовании новых типов специфичных документов – отредактировать инструкции в разделе <code>prompt_templates</code>, следуя правилам, указанным для соответствующего параметра в настоящей таблице
<code>title_page_min_length</code>	-	Минимальная длина титульной страницы. Необходимо установить значение, соответствующее минимальному количеству символов на титульной странице в используемых документах (применяется только для документов в формате <code>txt</code> и <code>rtf</code>)
<code>title_page_max_length</code>	-	Максимальная длина титульной страницы. Необходимо установить значение, соответствующее минимальному количеству символов на титульной странице в используемых документах (применяется только для документов в формате <code>txt</code> и <code>rtf</code>)
<code>links_section_templates</code>	-	Список шаблонов названий разделов со списками ссылок. Регистр вхождения в тексте не имеет значения
<code>digits_in_link_tpn_ignore_templates</code>	-	Список шаблонов для игнорирования проверки наличия цифры в <code>link_tpn</code>. Регистр вхождения в тексте не имеет значения

doc_types	-	Словарь для определения ссылок на документы. Необходимо добавить примеры в этот список, если Система не определяет их автоматически. Ключи словаря (ГОСТ, ОСТ, ПП – и.т.д) – извлекаемый из документа link_tpn (формальный тип документа, номер и дата). Значения словаря – список из возможных расшифровок краткого наименования формального типа документа. Для добавления нового типа документа необходимо добавить его в словарь, следуя указанному формату. Если у формального типа документа нет краткой формы – добавить обычную форму в качестве ключа и обязательно поставить двоеточие после названия ключа. Формат словаря: Сокращенное наименование НПА: - Полное наименование НПА (1 вариант) - Полное наименование НПА (2 вариант) НПА без полной формы 1: НПА без полной формы 2: Внимание! Данный словарь используется более чем в одном этапе извлечения сущностей, в настоящем разделе предоставлена ссылка на словарь, определяемый в разделе common_params: (см. Таблица 4-9)
-----------	---	--

ООО «ГК «ИННОТЕХ»
СИСТЕМА ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДОКУМЕНТОВ

not_links	-	Список вхождений, которые не должны быть извлечены как ссылки. Для добавления типа ссылки в исключения, которые извлекать не требуется, необходимо добавить новую строчку к существующему списку по правилу для атрибута links_npa_types
expected_outputs_first	-	Словарь с ожидаемым форматом вывода модели – первая часть. Не следует вносить в словарь настоящие ссылки на НПА; необходимо использовать несуществующие номера. Редактировать необходимо совместно с разделом prompt_templates: npa_extract_system_template: examples_instructions
expected_outputs_second	-	Словарь с ожидаемым форматом вывода модели – вторая часть. Не следует вносить в словарь настоящие ссылки на НПА; необходимо использовать несуществующие номера. Редактировать необходимо совместно с разделом prompt_templates: npa_extract_system_template: examples_instructions
additional_tpn_examples	-	Дополнительные примеры номеров в link_tpn. Редактировать необходимо совместно с разделом prompt_templates: npa_extract_system_template: general_instructions
prompt_templates	Содержит разделы:	Инструкции для модели с учетом наполнения списков

	– prompt_templates: npa_extract_title_system_template	Инструкция для модели, составлена на английском языке, содержит требование не извлекать из титульной страницы имя и тип документа как ссылку на другой документ. При изменении инструкции необходимо соблюдать следующие правила: – использовать английский язык; – соблюдать грамматические, лексические, синтаксические и этические нормы английского языка; – избегать формулировок, допускающих неоднозначное толкование; – оставить в тексте поле {} для последующей вставки примеров/текстов (с необходимым пояснением)
	– prompt_templates: npa_extract_title_system_template_document_name	Содержит разделы
	Содержит разделы:	
	– prompt_templates: npa_extract_system_template: general_instructions	Инструкция для модели, составлена на английском языке, содержит постановку задачи по выделению ссылок на НПА. При изменении инструкции необходимо соблюдать следующие правила: – использовать английский язык;

		– соблюдать грамматические, лексические, синтаксические и этические нормы английского языка; – избегать формулировок, допускающих неоднозначное толкование; – оставить в тексте поле {links_tpn_example}, {additional_tpn_examples_0}, {additional_tpn_examples_1}, {additional_tpn_examples_2} для последующей вставки примеров/текстов (с необходимым пояснением); – количество полей типа {additional_tpn_examples_N} должно соответствовать количеству элементов в поле additional_tpn_examples (как указано выше в настоящей таблице); – при изменении примеров в инструкции следует отредактировать также поле additional_tpn_examples
	– prompt_templates:npa_extract_system_template:important_notes	Инструкция для модели, составлена на английском языке, содержит важные замечания к постановке задачи по выделению ссылок на НПА. При изменении инструкции необходимо соблюдать следующие правила: – использовать английский язык; – соблюдать грамматические, лексические, синтаксические и этические нормы английского языка; – избегать формулировок, допускающих неоднозначное толкование;

		– не использовать более 12 пунктов с замечаниями; – оставить в тексте поле {doc_title_extraction} без пояснения и поле {not_links} (с необходимым пояснением)
	– prompt_templates: npa_extract _ system_template: examples_instructions	Инструкция для модели, составлена на английском языке, содержит примеры к постановке задачи по выделению ссылок на НПА. При изменении инструкции необходимо соблюдать следующие правила: – использовать английский язык; – соблюдать грамматические, лексические, синтаксические и этические нормы английского языка; – избегать формулировок, допускающих неоднозначное толкование; – оставить в тексте поле {expected_outputs_first} и {expected_outputs_second} (с необходимым пояснением); – при изменении примеров в инструкции следует отредактировать также поля expected_outputs_first и expected_outputs_second
	– prompt_templates: npa_extract_human_template	Инструкция для модели, составлена на английском языке, содержит дополнительную информацию и требования к ожидаемому формату вывода модели. Редактирование этой инструкции не допускается

ООО «ГК «ИННОТЕХ»
СИСТЕМА ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДОКУМЕНТОВ

input_tokens_limit	-	Максимальное количество токенов в чанке документа при последовательном просмотре документа. Не допускается указывать более 3000 токенов. Увеличение этого значения ускоряет извлечение ссылок из документа
filters: patterns_match	-	Список шаблонов регулярных выражений, используемых для фильтрации итогового списка ссылок на НПА. При соответствии вхождения ссылки любому из шаблонов из этого списка, такое вхождение не будет извлечено из документа
filters: max_link_tpn_length	-	Фильтр для исключения ссылок с длиной менее указанного числа. Не применяется к названию ссылочного документа
filters: min_link_tpn_length	-	Фильтр для исключения ссылок с длиной более указанного числа. Не применяется к названию ссылочного документа

Таблица 4-6 – Общие правила обучения Системы и изменяемые специальные атрибуты для выделения фактически существующих приложений в тексте документа и упоминаемых по тексту (подраздел config)

Параметр	Описание
Общие правила обучения Системы	Необходимо выполнить следующие шаги: – для дообучения Системы при использовании специфичных документов и получения релевантного результата отредактировать инструкцию для выделения некорректных кандидатов в заголовки приложений в параметре SYSTEM_PROMPT_TITLE_CORRECTOR и USER_PROMPT_TITLE_CORRECTOR, следуя правилам, указанным для соответствующего параметра в настоящей таблице; – для дообучения Системы при использовании специфичных документов и получения релевантного результата отредактировать инструкцию для выделения истинных ссылок на приложения в параметре SYSTEM_PROMPT_REF_CORRECTOR и USER_PROMPT_REF_CORRECTOR, следуя правилам, указанным для соответствующего параметра в настоящей таблице; – для дообучения Системы при использовании специфичных документов и получения релевантного результата отредактировать инструкцию для получения нормализованного написания заголовков приложений в параметре SYSTEM_PROMPT_HEADER_NORMALIZER и USER_PROMPT_HEADER_NORMALIZER, следуя правилам, указанным для соответствующего параметра в настоящей таблице; – для дообучения Системы при использовании специфичных документов и получения релевантного результата отредактировать инструкцию для подсчета количества вхождений ссылок на приложения в параметре SYSTEM_PROMPT_REF_COUNTER и USER_PROMPT_REF_COUNTER, следуя правилам, указанным для соответствующего параметра в настоящей таблице
prompts: SYSTEM_PROMPT_TITLE_CORRECTOR	Инструкция для модели, составлена на английском языке, содержит постановку задачи по идентификации кандидатов в заголовки приложений. При изменении инструкции необходимо соблюдать следующие правила: – использовать английский язык;

	– соблюдать грамматические, лексические, синтаксические и этические нормы английского языка; – избегать формулировок, допускающих неоднозначное толкование; – оставить в тексте поле {candidate} и {context} (с необходимым пояснением)
prompts: USER_PROMPT_REF_CORRECTER	Инструкция для модели, составлена на английском языке, содержит постановку задачи по идентификации кандидатов в заголовки приложений. При изменении инструкции необходимо соблюдать следующие правила: – использовать английский язык; – соблюдать грамматические, лексические, синтаксические и этические нормы английского языка; – избегать формулировок, допускающих неоднозначное толкование; – оставить в тексте поле {candidate} и {context} (с необходимым пояснением)
prompts: SYSTEM_PROMPT_REF_CORRECTER	Инструкция для модели, составлена на английском языке, содержит постановку задачи по идентификации истинных ссылок на приложения, определенных как некорректные. При изменении инструкции необходимо соблюдать следующие правила: – использовать английский язык; – соблюдать грамматические, лексические, синтаксические и этические нормы английского языка; – избегать формулировок, допускающих неоднозначное толкование; – оставить в тексте поле {candidate} и {context} (с необходимым пояснением)
prompts: USER_PROMPT_REF_CORRECTER	Инструкция для модели, составлена на английском языке, содержит постановку задачи по идентификации истинных ссылок на приложения, определенных как некорректные. При изменении инструкции необходимо соблюдать следующие правила: – использовать английский язык; – соблюдать грамматические, лексические, синтаксические и этические нормы английского языка; – избегать формулировок, допускающих неоднозначное толкование;

	– оставить в тексте поле {candidate} и {context} (с необходимым пояснением)
prompts: SYSTEM_PROMPT_HEADER_NORMALIZER	Инструкция для модели, составлена на английском языке, содержит постановку задачи по нормализации написания заголовков. При изменении инструкции необходимо соблюдать следующие правила: – использовать английский язык; – соблюдать грамматические, лексические, синтаксические и этические нормы английского языка; – избегать формулировок, допускающих неоднозначное толкование; – оставить в тексте поле {headers} (с необходимым пояснением)
prompts: USER_PROMPT_HEADER_NORMALIZER	Инструкция для модели, составлена на английском языке, содержит постановку задачи по нормализации написания заголовков. При изменении инструкции необходимо соблюдать следующие правила: – использовать английский язык; – соблюдать грамматические, лексические, синтаксические и этические нормы английского языка; – избегать формулировок, допускающих неоднозначное толкование; – оставить в тексте поле {headers} (с необходимым пояснением)
prompts: SYSTEM_PROMPT_REF_COUNTER	Инструкция для модели, составлена на английском языке, содержит постановку задачи по подсчету количества вхождений ссылок на приложения. При изменении инструкции необходимо соблюдать следующие правила: – использовать английский язык; – соблюдать грамматические, лексические, синтаксические и этические нормы английского языка; – избегать формулировок, допускающих неоднозначное толкование; – оставить в тексте поле {headers} и {references} (с необходимым пояснением)
prompts: USER_PROMPT_REF_COUNTER	Инструкция для модели, составлена на английском языке, содержит постановку задачи по подсчету количества вхождений ссылок на приложения. При изменении инструкции необходимо соблюдать следующие правила:

	– использовать английский язык; – соблюдать грамматические, лексические, синтаксические и этические нормы английского языка; – избегать формулировок, допускающих неоднозначное толкование; – оставить в тексте поле {headers} и {references} (с необходимым пояснением)
--	---

Таблица 4-7 – Общие правила обучения Системы и изменяемые специальные атрибуты для выделения наименований разделов документа (подраздел config)

Параметр	Описание
Общие правила обучения Системы	Необходимо выполнить следующие шаги: – при получении нерелевантного результата при использовании новых типов специфичных документов отредактировать параметры фильтрации результатов в разделе filters, следуя правилам, указанным для соответствующего параметра в настоящей таблице; – при получении нерелевантного результата при использовании новых типов специфичных документов отредактировать инструкции в разделе prompts: chapter_extraction_template, следуя правилам, указанным для соответствующего параметра в настоящей таблице
line_max_length	Максимальная длина строки. Строки длиннее этого значения будут разделены и восприняты Системой как независимые кандидаты в заголовки разделов документа
additional_config: context_size_head	Количество строк контекста для определения заголовка раздела документа (выше рассматриваемых кандидатов в заголовки)
additional_config: context_size_tile	Количество строк контекста для определения заголовка раздела документа (ниже рассматриваемых кандидатов в заголовки)
additional_config: num_lines_max	Максимальное количество строк в передаваемом в модель «окне» просмотра текста. Не следует устанавливать значение более чем 20. Для ускорения скорости извлечения заголовков раздела документа необходимо увеличить значения
filters: patterns_search	Список шаблонов для исключения кандидатов в заголовки разделов документа. Используется re.search без re.IGNORECASE

filters: patterns_search_ignorecase	Список шаблонов для исключения кандидатов в заголовки разделов документа. Используется re.search с re.IGNORECASE
filters: patterns_match	Список шаблонов для исключения кандидатов в заголовки разделов документа. Используется re.match
filters: max_header_length	Фильтр для исключения кандидатов в заголовки разделов документа с длиной более указанного числа
filters: min_header_length	Фильтр для исключения кандидатов в заголовки разделов документа с длиной менее указанного числа
prompts: cot_token	Специальный токен для маркировки служебных сообщений модели. Не допускается изменять
prompts: chapter_extraction_template: general_instructions	Инструкция для модели, составлена на английском языке, содержит постановку задачи по извлечению кандидатов в заголовки разделов документа. При изменении инструкции необходимо соблюдать следующие правила: – использовать английский язык; – соблюдать грамматические, лексические, синтаксические и этические нормы английского языка; – избегать формулировок, допускающих неоднозначное толкование
prompts: chapter_extraction_template: sample_section	Инструкция для модели, составлена на английском языке, содержит поле с передаваемым текстом и описание указанного поля. Не допускается редактирование этой инструкции
prompts: chapter_extraction_template: format_instructions	Инструкция для модели, составлена на английском языке, содержит дополнительную информацию и требования к ожидаемому формату вывода модели. Не допускается редактирование этой инструкции

Таблица 4-81 – Общие правила обучения Системы и изменяемые специальные атрибуты для выделения названия документа (подраздел config)

Параметр	Описание
Общие правила обучения Системы	Необходимо выполнить следующие шаги: – указать требуемые типы ссылок в параметре doc_types;

	– отредактировать указанный параметр в разделе <code>common_params</code> конфигурационного файла (см. Таблица 4-10); – указать необходимые исключения для типов документа в параметре <code>headers_FP</code>; – при получении нерелевантного результата при использовании новых типов специфичных документов отредактировать инструкции в разделе <code>template_base</code>, <code>template_type</code> и <code>template_tpn</code>, следуя правилам, указанным для соответствующего параметра в настоящей таблице. Необходимо обратить внимание на следующее: – <code>template_base</code> – используется для определения заголовка документа и формального типа + номера + даты документа, указанного в тексте; – <code>template_type</code> – используется для определения типа документа в соответствии со словарем <code>doc_types</code>; – <code>template_tpn</code> – используется для определения формального типа в сокращении (в соответствии со словарем <code>doc_types</code>) + номера + даты документа, указанного в тексте
<code>doc_types</code>	Словарь для определения типов документов. Необходимо добавить примеры в этот список, если Система не определяет их автоматически. Ключи словаря (ГОСТ, ОСТ, ПП – и.т.д) – краткое наименование (аббревиатура) типа документа. Значения словаря – список из возможных расшифровок краткого наименования типа документа. Для добавления нового типа документа необходимо добавить его в словарь, следуя указанному формату. Если у типа документа нет краткой формы – добавить обычную форму в качестве ключа и обязательно поставить двоеточие после названия ключа. Формат словаря: Сокращенное наименование типа документа: – Полное наименование (1 вариант) – Полное наименование (2 вариант) Тип документа без полной формы 1: Тип документа без полной формы 2: Внимание! Данный словарь используется более чем в одном этапе извлечения сущностей, в настоящем разделе предоставлена ссылка на словарь, определяемый в разделе <code>common_params</code>: (см. Таблица 4-10)

headers_FP	Примеры вхождений, которые не должны быть определены как тип или наименование документа
template_base: general_instructions	Инструкция для модели, составлена на английском языке, содержит постановку задачи по выделению названия документа, а также типа и номера документа. При изменении инструкции необходимо соблюдать следующие правила: – использовать английский язык; – соблюдать грамматические, лексические, синтаксические и этические нормы английского языка; – избегать формулировок, допускающих неоднозначное толкование
template_base: important_notes	Инструкция для модели, составлена на английском языке, содержит содержит важные замечания к постановке задачи по выделению названия документа, а также типа и номера документа. При изменении инструкции необходимо соблюдать следующие правила: – использовать английский язык; – соблюдать грамматические, лексические, синтаксические и этические нормы английского языка; – избегать формулировок, допускающих неоднозначное толкование
template_base: format_instructions	Инструкция для модели, составлена на английском языке, содержит дополнительную информацию и требования к ожидаемому формату вывода модели. Не допускается редактирование этой инструкции
template_base: sample_section	Инструкция для модели, составлена на английском языке, содержит поле с передаваемым текстом и описание указанного поля. Не допускается редактирование этой инструкции
template_type: general_instructions	Инструкция для модели, составлена на английском языке, содержит постановку задачи по выделению типа документа (краткого наименования типа). При изменении инструкции необходимо соблюдать следующие правила: – использовать английский язык; – соблюдать грамматические, лексические, синтаксические и этические нормы английского языка; – избегать формулировок, допускающих неоднозначное толкование

template_type: important_notes	Инструкция для модели, составлена на английском языке, содержит важные замечания к постановке задачи по выделению типа документа (краткого наименования типа). При изменении инструкции необходимо соблюдать следующие правила: – использовать английский язык; – соблюдать грамматические, лексические, синтаксические и этические нормы английского языка; – избегать формулировок, допускающих неоднозначное толкование
template_type: format_instructions	Инструкция для модели, составлена на английском языке, содержит дополнительную информацию и требования к ожидаемому формату вывода модели. Не допускается редактирование этой инструкции
template_type: sample_section	Инструкция для модели, составлена на английском языке, содержит поле с передаваемым текстом и описание указанного поля. Не допускается редактирование этой инструкции
template_tpn: general_instructions	Инструкция для модели, составлена на английском языке, содержит постановку задачи по выделению нормализованного типа и номера документа. При изменении инструкции необходимо соблюдать следующие правила: – использовать английский язык; – соблюдать грамматические, лексические, синтаксические и этические нормы английского языка; – избегать формулировок, допускающих неоднозначное толкование
template_tpn: important_notes	Инструкция для модели, составлена на английском языке, содержит важные замечания к постановке задачи по выделению нормализованного типа и номера документа. При изменении инструкции необходимо соблюдать следующие правила: – использовать английский язык; – соблюдать грамматические, лексические, синтаксические и этические нормы английского языка; – избегать формулировок, допускающих неоднозначное толкование

template_tpn: format_instructions	Инструкция для модели, составлена на английском языке, содержит дополнительную информацию и требования к ожидаемому формату вывода модели. Не допускается редактирование этой инструкции
template_tpn: sample_section	Инструкция для модели, составлена на английском языке, содержит поле с передаваемым текстом и описание указанного поля. Не допускается редактирование этой инструкции.
template_correction	Инструкция для модели, составлена на английском языке, содержит информацию и требования к исправлению некорректного вывода модели. Не допускается редактирование этой инструкции
title_page_min_length	Минимальная длина титульной страницы. Необходимо установить значение, соответствующее минимальному количеству символов на титульной странице в используемых документах (применяется только для документов в формате txt и rtf)
title_page_max_length	Максимальная длина титульной страницы. Необходимо установить значение, соответствующее минимальному количеству символов на титульной странице в используемых документах (применяется только для документов в формате txt и rtf)

Таблица 4-92 – Общие правила обучения Системы и изменяемые специальные атрибуты для определения принадлежности к установленным типам документов (подраздел config)

Параметр	Описание
Общие правила обучения Системы	Необходимо выполнить следующие шаги: – при получении нерелевантного результата при использовании новых типов специфичных документов отредактировать список с ключевыми образцами текстов для одного из определяемых флагов принадлежности к установленным типам документов; – допускается редактирование вхождений текстов, добавление и удаление вхождений текстов; – не допускается установка новых ключей (разделов) в словарь pattern_dict; – не допускается отсутствие списка вхождений текстов в любом из ключей словаря pattern_dict
title_page_min_length	Минимальная длина титульной страницы. Необходимо установить значение, соответствующее минимальному количеству символов на титульной

	странице в используемых документах (применяется только для документов в формате txt и rtf)
title_page_max_length	Максимальная длина титульной страницы. Необходимо установить значение, соответствующее минимальному количеству символов на титульной странице в используемых документах (применяется только для документов в формате txt и rtf)
pattern_dict:	Словарь с масками для извлечения флагов типов документа. Добавление новых ключей в словарь не допускается. Должны присутствовать только ключи: – rosreestr_flg; – egrul_flg; – egrip_flg; – work_safety_doc_flg; – order_dismissal_doc_flg; – work_records_doc_flg. Добавление новых флагов типов документа выполняется только разработчиком Системы. Значения в списках в ключах могут быть отредактированы, а также могут быть добавлены новые значения. Система не учитывает окончания и склонение слов при определении флага
pattern_dict: rosreestr_flg	Список ключевых образцов текста для отнесения документа к одному из типов, определяемому в проверке (Росреестр)
pattern_dict: egrul_flg	Список ключевых образцов текста для отнесения документа к одному из типов, определяемому в проверке (Единый государственный реестр юридических лиц)
pattern_dict: egrip_flg	Список ключевых образцов текста для отнесения документа к одному из типов, определяемому в проверке (Единый государственный реестр индивидуальных предпринимателей)
pattern_dict: work_safety_doc_flg	Список ключевых образцов текста для отнесения документа к одному из типов, определяемому в проверке (Инструктаж по охране труда)
pattern_dict: order_dismissal_doc_flg	Список ключевых образцов текста для отнесения документа к одному из типов, определяемому в проверке (Приказ на увольнение)
pattern_dict: work_records_doc_flg	Список ключевых образцов текста для отнесения документа к одному из типов, определяемому в проверке (Трудовая книжка)

В конфигурационном файле содержится информация о подключениях к БЯМ и БД Реестра документов Системы.

Разделы и атрибуты подключений приведены в **Таблице 4-9**.

Таблица 4-103 – Изменяемые параметры вне раздела *attributes*

Раздел	Параметр	Описание
common_params:	component_name	Имя компонента. Не допускается изменять
common_params:	incoming_event_type	Имя входящего события, на которое должен реагировать компонент. Не допускается изменять
common_params:	outcoming_event_type	Имя исходящего события, которое должен создавать компонент. Не допускается изменять
common_params:	component_endpoint	Адрес эндпойнта с реализацией основной функции компонента. Должен соответствовать разделу common_params: uvicorn
common_params: llm_connection	-	Параметры для подключения к LLM
-	model	Путь до файлов модели
-	openai_api_base	Адрес хоста с моделью LLMs
common_params: external_db: connection:	-	Параметры подключения к БД Реестра документов Системы
-	drivername	Системное имя драйвера, используемого для подключения к БД (postgresql для подключения к БД PostgreSQL)
-	host	URL сервера СУБД (пример: 10.10.165.20)
-	port	Порт сервера СУБД (пример: 5432)

-	database	Системное имя БД Реестра документов Системы
common_params: uvicorn:	-	Параметры внутреннего сервера uvicorn
-	host	Имя хоста для внутреннего сервера uvicorn (пример: localhost)
-	port	Номер порта для внутреннего сервера uvicorn (пример: 9001)
-	workers	Количество рабочих процессов uvicorn. Не допускается изменять
-	log_config	Специальные параметры логгирования uvicorn. Не допускается изменять
common_params: call_endpoint	-	Специальные параметры для функции вызова эндпойнтов внутреннего сервера uvicorn. Не допускается изменять
-	module_class_name	Имя модуля, реализующего вызов. Не допускается изменять
-	additional_params: fastapi_initializer	Приложение fastapi. Не допускается изменять
common_params: logger	-	Параметры логирования работы компонента
-	log_file	Путь до файла лога (внутри контейнера)
-	log_format	Формат записей в файле лога
dynamical_config_loading	use	Специальный параметр – флаг, требующий динамического обновления конфигурации во время работы Системы

common_params:	doc_types	Словарь типов документов. Используется в виде ссылки в одноименных параметрах из Таблиц 4-5 и 4-8 настоящего Руководства
----------------	-----------	--

4.8. Работа со справочником обязательных разделов документа

Справочник обязательных разделов документа (Справочник) размещается в файле `usercheck_config.yml`, который расположен по адресу: `EXT_CONFIG_PATH=/config/usercheck/usercheck_config.yml`.

Для изменения Справочника следует загрузить новый (скорректированный) конфигурационный файл в директорию `/config/usercheck/` на сервере FTP любым из инструментов сетевого копирования файлов. Дополнительно перезапускать сервис не требуется.

Справочник имеет двухуровневую структуру:

- первый уровень – тип документа (атрибут `name`);
- второй уровень – список обязательных разделов для соответствующего типа документа (атрибут `sections.mandatory`).

Структура данных Справочника в формате `yml` следующая:

```
usercheck:
  checks:
    - id: 5
      doc_types:
        - name: "ГОСТ"
          sections:
            mandatory:
              - "Предисловие"
              - "Термины и определения"
            optional:
        - name: "ГОСТ Р"
          sections:
            mandatory:
              - "Предисловия"
              - "Термины и определения"
              - "Область применения"
            optional:
        - name: "ГОСТ РВ"
          sections:
            mandatory:
              - "Предисловия"
              - "Термины и определения"
              - "Область применения"
              - "Нормативные ссылки"
            optional:
```

При поставке Системы Справочник содержит списки обязательных разделов для следующих 15 типов документов:

- Межгосударственный стандарт (ГОСТ);

- Национальный стандарт Российской Федерации (ГОСТ Р);
- Государственные стандарты РФ в области обороны и военной промышленности (ГОСТ РВ);
- Национальный стандарт РФ, гармонизированный с ИСО и МЭК (ГОСТ Р ИСО / МЭК);
- Отраслевой стандарт (ОСТ);
- Технические условия (ТУ);
- Стандарт организации (СТО);
- Закон Российской Федерации;
- Федеральный закон;
- Постановление Правительства РФ;
- Распоряжение Правительства РФ;
- Приказ Минэкономразвития России;
- Приказ Министерства образования и науки Российской Федерации;
- Указ Президента Российской Федерации;
- Кодекс Российской Федерации;
- Единый кодификатор предметов снабжения для федеральных государственных нужд.

Для добавления списка обязательных разделов нового типа документов необходимо выполнить следующие шаги:

- a) Создать новый элемент с атрибутом name.
- b) На втором атрибутивном уровне нового элемента дополнить необходимый список значений в параметр sections.mandatory.

Для редактирования списка обязательных разделов существующего типа документов необходимо выполнить следующие шаги:

- a) Найти соответствующий элемент по значению name.
- b) Внести требуемые изменения в атрибут sections.mandatory.

5. Планово-предупредительные мероприятия

Представленные процедуры предназначены для регулярной проверки работоспособности системы «Система интеллектуального анализа документов» и локализации неработающих компонентов. Регулярное проведение данных мероприятий (см. **Таблица 5-1**) позволяет своевременно выявлять и устранять неполадки, обеспечивая стабильную работу системы.

Таблица 5-1 — Перечень планово-предупредительных мероприятий

№	Содержание мероприятия	Периодичность	Ответственный
1	Мониторинг работоспособности различных компонентов системы	Определяется Администратором (рекомендованная частота: один раз в сутки, 5 дней в неделю)	Администратор
2	Локализация неработающих компонентов системы	Определяется Администратором	Администратор
3	Проверка наличия сертификатов	Перед запуском пилотной или целевой версии системы в промышленную среду	Администратор
4	Проверка в консоли администратора Kubernetes состояния подов (все поды запущены в требуемом количестве)	После запуска пилотной или целевой версии системы в промышленную среду	Администратор
5	Проверка лог-файлов запуска компонентов продукта (микросервисов) на наличие сообщений с типом «ERROR»	После запуска пилотной или целевой версии системы в промышленную среду	Администратор
6	Проверка работоспособности микросервисов на маршрутах (/healthcheck)	После запуска пилотной или целевой версии системы в промышленную среду	Администратор
7	Проверка работоспособности микросервисов	Ежедневно	Администратор
8	Проверка работоспособности компонента ArangoDB	Ежедневно	Администратор
9	Проверка работоспособности компонента БД PostgreSQL	Ежедневно	Администратор
10	Проверка статуса запуска серверов	Ежедневно	Администратор
11	Проверка статуса мониторинга использования серверов	Раз в месяц	Администратор
12	Проверка статуса мониторинга выполнения задач планировщика	Ежедневно	Администратор

№	Содержание мероприятия	Периодичность	Ответственный
13	Мониторинг использования серверов: анализ статистики по лог-файлам	Раз в месяц	Администратор
14	Резервное копирование	Один раз в сутки в период с 22.00 до 08.00 по МСК, 5 дней в неделю	Администратор

5.1. Документирование результатов

Порядок документирования результатов:

- Все результаты проверки работоспособности системы и локализации неработающих компонентов должны быть задокументированы;
- Документация должна содержать информацию о дате проверки, выявленных проблемах, принятых мерах и результате их применения;
- Документация должна храниться в системе контроля версий или в специализированной системе управления инцидентами;
- Это позволит отследить историю работы платформы и легко идентифицировать причины возникновения ошибок.

6. Аварийное восстановление системы «Система интеллектуального анализа документов»

Аварийно-восстановительные работы проводятся с целью поддержки бесперебойной работы системы и восстановления работоспособности системы в максимально сжатые сроки без потери данных.

6.1. Критерии аварийного события

1. Недоступность приложения «Система интеллектуального анализа документов».
2. Отказ одной из частей ИС.
3. Массовые обращения пользователей системы в службу поддержки.
4. Выявление неисправности в результате планово-предупредительных работ со стороны службы поддержки.
5. Обращение в службу поддержки со стороны сотрудников мониторинга, администраторов СУБД и серверов приложений.
6. Не пройдены проверки работоспособности при вводе системы в эксплуатацию.

К потенциальным аварийным событиям также можно отнести:

1. Недоступность серверов приложений.
2. Недоступность баз данных.
3. Некорректная работа системы.

6.2. Порядок действий при выявлении аварийного события

Общий порядок действий при возникновении аварийного события:

1. Диагностика возникшей ситуации, первичный сбор информации.
2. Оповещение координатора, ответственного за выполнение аварийного восстановления, если самостоятельное устранение проблемы невозможно.
3. Определение специфики проблемы, привлечение специалистов.
4. Выполнение действий по восстановлению согласно выбранному сценарию.
5. Рассылка оповещений об аварийной ситуации:
 - всем заинтересованным лицам (например, Бизнес-заказчик, Аналитик ИС);
 - каналы оповещения выбираются на усмотрение службы поддержки в зависимости от критичности ситуации (например, электронная почта, SMS-сообщение, звонок);
 - в оповещении указывается следующая информация: данные о зарегистрированном инциденте (описание аварийного события, периметр события, возможные причины возникновения, плановые сроки устранения, влияние на смежные ИС и возможные последствия).

К процессам поддержки и восстановления работоспособности системы должны быть привлечены следующие сотрудники:

- Системные администраторы;
- Администраторы серверов приложений (в части сопровождения и эксплуатации систем поддержки бизнеса).

6.3. Общие рекомендации по восстановлению деятельности системы «Система интеллектуального анализа документов»

1. Проверить статус работоспособности приложения, при необходимости, перезапустить.
2. Закончилось место на диске - необходимо освободить место при помощи архивирования и удаления данных с дисков.
3. Истек срок действия пароля для ТУЗ - необходимо актуализировать пароль для ТУЗ.
4. Сломался сервер с виртуализацией - необходимо проанализировать ситуацию и устранить проблему.