

ООО «Иксими»

**Описание функциональных характеристик
программного обеспечения
«Интеллектуальная корпоративная система
интеграционного менеджмента информации
"ИКСИМИ:ДОКУМЕНТЫ 5.0"»**

г. Зеленодольск

2026 год

Настоящий документ содержит описание функциональных характеристик программного обеспечения «Интеллектуальная корпоративная система интеграционного менеджмента информации "ИКСИМИ:ДОКУМЕНТЫ 5.0"»

Назначение и область применения

Программное обеспечение «Интеллектуальная корпоративная система интеграционного менеджмента информации "ИКСИМИ:ДОКУМЕНТЫ 5.0"» применяется при обработке информации содержащейся в документах, для поиска, анализа и управления данными, представленными в базах знаний, сформированных на основе документов организации.

Программа предназначена для централизованной работы с информацией и обеспечения поиска документов и информации в них по ключевым словам, по смыслу, а также получения ответов на вопросы пользователя с использованием искусственного интеллекта, основанных на загруженных документах, используя RAG технологию.

Функциональные задачи

Платформа обеспечивает решение следующих задач:

Задачи интеллектуального поиска и генерации ответов

Платформа обеспечивает формирование релевантных ответов на пользовательские запросы строго на основе локальных источников информации, загруженных в базу знаний, с применением технологии Retrieval-Augmented Generation (RAG). Реализован гибридный механизм поиска, сочетающий семантический (sentence-transformers) и ключевой (TF-IDF) подходы, что гарантирует высокую точность и релевантность выдаваемых результатов. Система выполняет предобработку и обогащение пользовательских запросов, генерируя уточненные промежуточные запросы для повышения полноты извлекаемого контекста. Поддерживается двухстадийная модель извлечения данных: поиск может осуществляться как по всему пулу документов, так и в рамках конкретного документа, выбранного пользователем. Платформа обеспечивает возможность кросс-языковых семантических запросов на русском языке по мультязычным документам, что позволяет находить релевантную информацию независимо от языка оригинала.

Задачи загрузки, обработки и управления документами

Платформа обеспечивает загрузку PDF-документов с реализацией сценария автоматической обработки импортируемых данных. Выполняется семантическая фрагментация и векторизация загружаемых документов с формированием векторного хранилища базы знаний. Предусмотрена возможность обогащения метаданных документов импортом атрибутов из внешних источников в CSV-формате, что позволяет сформировать полноценную справочно-информационную систему. По каждому сгенерированному ответу система отображает список релевантных документов, на основе которых был построен ответ, обеспечивая контекстную точность и прозрачность работы ИИ.

Задачи навигации и визуализации контекста

Платформа обеспечивает подсветку базового фрагмента в тексте документа-источника, использованного ИИ в качестве контекста, с возможностью одним кликом перейти к данному фрагменту прямо из ответа чата. Это позволяет пользователю оперативно верифицировать информацию и глубже изучать первоисточники.

Задачи работы с текстом и AI-ассистентом документа

Платформа предоставляет инструменты работы с выделением текста. Пользователи могут выделять фрагменты текста в открытом документе и запрашивать у ИИ перевод выделенного текста, объяснение сложных терминов или концепций, а также краткое изложение содержания. Реализована возможность задавать вопросы, относящиеся к содержимому конкретного открытого документа, получая точные и контекстные ответы. После работы с выбранным документом пользователь может вернуться к общей базе знаний и продолжить задавать вопросы, охватывающие все доступные источники, что обеспечивает гибкость в работе с базой документов.

Задачи пользовательского интерфейса и взаимодействия

Платформа обеспечивает единый интерфейс документа с чатом, интегрируя окно просмотра и чат ИИ-ассистента в одном пространстве без необходимости переключения между вкладками. Поддерживается потоковый (stream) режим обработки и отображения данных в ИИ-чате, что ускоряет визуальное восприятие ответов и повышает комфорт взаимодействия. Предусмотрена возможность самостоятельной регистрации пользователей через встроенный интерфейс регистрации ПО.

Задачи администрирования и управления доступом

Платформа предоставляет панель администрирования с функциями просмотра и редактирования учетных записей пользователей. Обеспечивается гибкая настройка прав пользователей как на уровне интерфейса, так и на уровне RAG-специфичных прав, что позволяет с помощью инструментов администрирования определять права доступа пользователя к необходимым базам знаний. Это гарантирует разграничение доступа в соответствии с корпоративной политикой безопасности.

Задачи автономной работы и информационной безопасности

Платформа обеспечивает полностью автономную работу системы без подключения к сети Интернет в закрытом информационном контуре организации. Вся обработка, хранение, извлечение и ИИ-генерация выполняются локально на собственных мощностях предприятия, что гарантирует сохранность конфиденциальных данных и соответствие требованиям информационной безопасности.

Функциональные возможности

Интеллектуальный поиск и генерация ответов (RAG)

Система использует передовую технологию **Retrieval-Augmented Generation (RAG)** для работы с корпоративными данными:

- **Формирование ответов ИИ:** Генерация релевантных ответов на пользовательские запросы строго на основе локальных источников информации (загруженных документов), что исключает галлюцинации модели.
- **Гибридный поиск:** Комбинация семантического поиска (sentence-transformers) и ключевого поиска (TF-IDF) для максимальной точности и релевантности результатов.
- **Двухстадийная модель извлечения:** Поиск данных осуществляется либо по всему пулу документов, либо в рамках конкретного документа, выбранного пользователем.
- **Предобработка запросов:** Обогащение и уточнение промежуточных запросов перед поиском для повышения качества извлекаемого контекста.
- **Кросс-языковая поддержка:** Возможность задавать запросы на русском языке по мультязычным документам (семантический кросс-языковой подход).

Работа с документами и источниками

Инструменты для загрузки, обработки и визуализации содержимого базы знаний:

- **Загрузка и обработка:** Поддержка загрузки PDF-документов с автоматической семантической фрагментацией и векторизацией для формирования векторного хранилища.
- **Отображение источников:** По каждому запросу система показывает список релевантных документов, на основе которых был сгенерирован ответ.
- **Интерактивная навигация:** Подсветка базового фрагмента в тексте документа-источника, использованного ИИ в качестве контекста. Переход к этому фрагменту выполняется одним кликом из ответа чата.
- **Обогащение метаданных:** Возможность импорта атрибутов документов из внешних CSV-файлов для формирования полноценной справочно-информационной системы.
- **Единый интерфейс:** Работа с документом и чат-ассистентом интегрированы в одном окне, без необходимости переключения между вкладками.

Инструменты работы с текстом (AI-ассистент документа)

Специализированные функции при работе с открытым документом:

- **Работа с выделением:** Пользователь может выделить фрагмент текста и запросить у ИИ:
 - Перевод выделенного текста.
 - Объяснение сложных терминов или концепций.
 - Краткое изложение содержания.
- **Вопросы по документу:** Возможность задавать контекстные вопросы, относящиеся к содержимому конкретного открытого документа.
- **Гибкость переключения:** После завершения работы с выбранным документом пользователь может вернуться к общей базе знаний и продолжить работу со всеми доступными источниками.

Пользовательский интерфейс и взаимодействие

Удобство работы и современные форматы отображения данных:

- **Потоковый режим (Stream):** Поддержка потокового режима обработки и отображения данных в ИИ-чате, что ускоряет визуальное восприятие ответов.
- **Интерфейс регистрации:** Самостоятельная регистрация новых пользователей через встроенный интерфейс ПО.

Администрирование и безопасность

Инструменты управления системой и разграничения доступа в закрытом контуре:

- **Панель администрирования:** Просмотр и редактирование учетных записей пользователей.
- **Гибкая настройка прав:** Определение прав доступа пользователей как на уровне интерфейса, так и на уровне RAG-специфичных прав (доступ к конкретным базам знаний).
- **Автономная работа:** Система функционирует полностью локально, без подключения к сети Интернет, в закрытом информационном контуре организации. Вся обработка, хранение, извлечение и ИИ-генерация выполняются на собственных мощностях предприятия.

Входные данные

Входными данными для ПО являются:

- Файл(ы) в формате PDF, CSV (опционально);
- Конфигурационные файлы, определяющие конфигурацию базы знаний и параметры используемых моделей (на базе заполненной формы формирования базы);
- Пользовательские запросы в веб-интерфейсе.

* Предоставленный для экспертной проверки экземпляр ПО включает демонстрационный корпус из 5 тестовых электронных документов, уже прошедших полную автоматическую обработку и сформированную конфигурацию. Запросы к тестовой базе знаний доступны в веб-интерфейсе сразу после развертывания ПО

Выходные данные

По результатам обработки электронных документов ПО формирует следующие выходные данные:

- База знаний, включающая: конфигурацию, фрагменты, установки, векторный индекс, пользовательские права

По результатам обработки запроса ПО формирует следующие выходные данные:

- Список релевантных документов
- Ответ ИИ Ассистента
- Указатель на источник информации

Результаты обработки доступны через веб-интерфейс Платформы

Архитектура ПО

Программное обеспечение «Интеллектуальная корпоративная система интеграционного менеджмента информации "ИКСИМИ:ДОКУМЕНТЫ 5.0"» состоит из следующих частей:

- 1) Базовые библиотеки
 - Flask

- Flask-cors
 - Werkzeug
- 2) Работа с БД PostgreSQL
- Psycopg2-binary
- 3) Обработка PDF и изображений
- Pfdplumber
 - Pfdminer.six
 - Pillow
- 4) Обработка текста и NLP
- Spacy
 - Sentence-transformers
 - Scikit-learn
 - Pymorphy2
- 5) Работа с данными
- Pandas
- 6) API AI
- Openai
- 7) Прочее
- Requests
 - Argparse

Взаимодействие между компонентами

Взаимодействие между компонентами осуществляется следующим образом.

1. Пользователь формирует запрос через веб-интерфейс.
2. Запрос передаётся в Backend посредством REST API
3. Backend выполняет:
 - предобработку запроса;
 - гибридный поиск;
 - формирование контекста.
4. Контекст передаётся языковой модели.
5. Полученный ответ возвращается пользователю.
6. При необходимости пользователь открывает документ, а система отображает стартовый использованный фрагмент с использованием координат, полученных на этапе обработки PDF.

Пайплайн обработки документов включает следующие стадии:

Extraction → Chunking → Embedding → Indexing → Retrieval → Ranking → Generation

Все стадии являются независимыми и могут заменяться без изменения архитектуры остальных компонентов.

Требования к техническим средствам (Эксплуатационные характеристики)

Для установки и эксплуатации программного обеспечения «Интеллектуальная корпоративная система интеграционного менеджмента информации

"ИКСИМИ:ДОКУМЕНТЫ 5.0"» необходимо, чтобы аппаратное обеспечение соответствовало следующим требованиям:

- ЦП: x86_64 архитектура, от 4 ядер
- ОЗУ: от 8 ГБ (зависит от объема индексируемой базы знаний и размера загруженной языковой модели).
- GPU: Опционально, рекомендуется NVIDIA с поддержкой CUDA 11.8+
- Дисковая подсистема: SSD от 20 ГБ (зависит от объема индексируемой базы знаний и размера загруженной языковой модели).
- ОС: Рекомендуется Astra Linux, ОСнова, или любой совместимый дистрибутив Linux
- Подключение к интернету (не требуется для работы ПО; может использоваться только на этапе установки для загрузки ПО опционально), 10-100 Мбит\с
- Интернет-браузеры – Яндекс.Браузер, Google Chrome, другие браузеры на движке Chromium

Описание функциональной части программного обеспечения

- Языки программирования программного обеспечения «Интеллектуальная корпоративная система интеграционного менеджмента информации "ИКСИМИ:ДОКУМЕНТЫ 5.0" », – JavaScript, Python
- СУБД программного обеспечения «Интеллектуальная корпоративная система интеграционного менеджмента информации "ИКСИМИ:ДОКУМЕНТЫ 5.0"» (Система управления векторными данными) - FAISS v1.14.x

Информация, необходимая для установки и эксплуатации

Для установки и правильной эксплуатации «Интеллектуальная корпоративная система интеграционного менеджмента информации "ИКСИМИ:ДОКУМЕНТЫ 5.0"» необходимо ознакомиться с информацией, размещенной в Руководстве пользователя.