

**ПРОГРАММНАЯ ПЛАТФОРМА ДЛЯ УПРАВЛЕНИЯ
КОРПОРАТИВНЫМИ ЗНАНИЯМИ, AI-
АССИСТЕНТАМИ И ИНТЕЛЛЕКТУАЛЬНЫМИ
АГЕНТАМИ «SEMANTIX»**

Функциональные характеристики ПО

Листов 14

АННОТАЦИЯ

В документе приведены сведения о программной платформе для управления корпоративными знаниями, AI-ассистентами и интеллектуальными агентами «Semantix» (далее – Semantix).

Настоящий документ содержит описание назначения, функциональных возможностей и решаемых Semantix задач.

СОДЕРЖАНИЕ

1. НАЗНАЧЕНИЕ ПРОГРАММЫ	4
1.1. Назначение Semantix	4
1.2. Основные характеристики	5
2. ФУНКЦИОНАЛЬНЫЕ ХАРАКТЕРИСТИКИ	6
2.1. Модуль организационных контуров и управления пользователями	6
2.2. Модуль выбора и подключения провайдеров AI-моделей	7
2.3. Модуль наборов данных, баз знаний и обработки документов	8
2.4. Модуль поиска, RAG и проверяемых ответов	8
2.5. Модуль AI-ассистентов и диалоговых приложений	9
2.6. Модуль AI-агентов и визуального конструктора	9
2.7. Мультиагентная логика и инструменты агента	10
2.8. Модуль коннекторов и синхронизации данных	10
2.9. Модуль MCP-серверов и внешних инструментов	11
2.10. Модуль API, SDK и внешней интеграции	12
ПЕРЕЧЕНЬ СОКРАЩЕНИЙ	13

1. НАЗНАЧЕНИЕ ПРОГРАММЫ

1.1. Назначение Semantix

Программная платформа Semantix предназначена для интеллектуальной обработки корпоративных данных, построения баз знаний, организации семантического поиска, создания и управления AI-ассистентами и AI-агентами (мультиагентами), использующими подключаемые языковые модели и внешние инструменты.

Платформа собирает знания из разнородных источников, формирует единый поисковый индекс, позволяет создавать проверяемые ответы со ссылками на первоисточники, настраивать AI-ассистентов и строить агентные сценарии обработки данных.

Semantix не является отдельной большой языковой моделью и не привязывает Заказчика к одному поставщику AI-моделей. Платформа выступает управляющим и оркестрационным слоем: она подключает внешние или локальные модели, управляет наборами данных, поиском, ассистентами, агентами, коннекторами, MCP-инструментами, правами доступа и журналами выполнения.

Semantix даёт пользователю возможности:

- формировать единое пространство корпоративных знаний из документов, файловых хранилищ, корпоративных систем, задач, сообщений, почты и иных источников разнородных данных;
- создавать наборы данных и базы знаний для последующего поиска, анализа и генерации ответов с указанием первоисточников;
- создавать и настраивать AI-ассистентов для консультаций, поиска, анализа документов и сопровождения внутренних или внешних пользователей;
- проектировать и создавать AI-агентные сценарии на Canvas, где задача выполняется цепочкой компонентов, агентов и инструментов;
- подключать разные AI-модели и провайдеры моделей без жесткой зависимости от одного поставщика;
- интегрироваться с внешними системами через коннекторы, API и MCP-серверы;
- разграничивать доступ пользователей по организационным контурам, ролям и объектам системы;
- получать журналируемые и проверяемые результаты выполнения поиска, генерации и агентных действий.

Основной практический результат применения Semantix – снижение трудоемкости поиска и анализа информации, повышение воспроизводимости ответов, ускорение обработки документов и обращений, а также создание и внедрения AI-агентов в корпоративный контур.

Semantix позволяет наиболее эффективно использовать даже небольшие AI-модели, достигая нужного результата. Таким образом, стоимость внедрения AI-решений в организации существенно снижается без потери качества результата.

1.2. Основные характеристики

В составе Semantix предусмотрены два взаимосвязанных, но функционально различных механизма интеллектуального взаимодействия: AI-ассистенты и AI-агенты.

AI-ассистент является основной точкой коммуникации пользователя с продуктом. Он обеспечивает диалоговый интерфейс, принимает запросы пользователя, обращается к подключенным наборам данных и базам знаний, формирует ответы на основании корпоративных данных и, при необходимости, инициирует выполнение специализированных агентов. AI-ассистент используется как управляемый канал взаимодействия с системой: пользователь задает вопрос или даёт поручение, а AI-ассистент определяет, достаточно ли поиска по знаниям, требуется ли обращение к внешнему инструменту или необходимо запустить агентный сценарий.

AI-агент представляет собой специализированный исполнительный компонент, предназначенный для автоматизации конкретного бизнес-процесса или отдельной прикладной задачи. AI-агент может быть собран в визуальной среде-конструкторе из набора функциональных блоков, включая получение входных данных, поиск по знаниям, вызов моделей, выполнение условий, обращение к инструментам, МСР-серверам и внешним системам. В отличие от AI-ассистента, AI-агент ориентирован не только на ведение диалога, но и на выполнение формализованной последовательности действий: анализ документа, обработку заявки, классификацию обращения, подготовку отчета, запуск интеграционного сценария или иное действие в рамках заданного процесса.

Таким образом, AI-ассистент выступает пользовательской и диалоговой точкой входа, а AI-агенты выступают исполняемыми сценариями автоматизации, которые могут вызываться AI-ассистентом, пользователем или внутренними механизмами Semantix в зависимости от настроек конкретного решения.

В состав Semantix входят базовые компоненты, указанные в таблице.

Компонент	Назначение	Основной сценарий использования	Результат работы
AI-ассистент	Диалоговая точка взаимодействия пользователя с Semantix	Пользователь задает вопрос, запрашивает сведения, просит выполнить действие или получить ответ по корпоративным данным	Ответ пользователю, ссылка на источники, запуск агента или инструмента при необходимости
AI-агент	Исполнительный сценарий для решения конкретной задачи или бизнес-процесса	Обработка документа, анализ обращения, классификация, подготовка отчета, выполнение цепочки действий через Canvas	Выполненная задача, сформированный результат, отчет, действие во внешней системе
Конструктор БП	Визуальная среда проектирования агентных сценариев	Настройка логики агента из блоков, условий, моделей, инструментов и источников данных	Исполняемый агентный процесс
Набор данных / база знаний	Источник проверенных корпоративных знаний для ассистентов и агентов	Поиск релевантных фрагментов, обогащение ответа, RAG-сценарии	Контекст, фрагменты документов, ссылки на первоисточники
МСР / инструменты	Механизм подключения внешних функций и систем к агентам	Вызов внешнего сервиса, API, корпоративной системы или специализированного инструмента	Данные или действие, полученные из внешней системы

2. ФУНКЦИОНАЛЬНЫЕ ХАРАКТЕРИСТИКИ

2.1. Модуль организационных контуров и управления пользователями

Организационный контур (рабочее пространство организации) – термин, введенный в пользовательской и технической документации Semantix вместо англоязычного термина tenant.

Модуль обеспечивает многопользовательскую и многоорганизационную работу:

- создание организационного контура администратором системы;
- хранение имени контура и набора моделей по умолчанию: LLM, embedding, ASR, image-to-text, rerank, TTS и сервисная модель;
- создание корневой папки и инициализация набора моделей для нового контура;

- добавление пользователей в контур по email или по идентификатору существующей учетной записи;
- создание нового пользователя при добавлении в контур при наличии соответствующего права;
- назначение ролей владельца, администратора и обычного пользователя;
- запрет добавления суперпользователей в обычные контуры;
- получение списка пользователей контура, списка всех контуров и списка всех пользователей с их ролями;
- удаление пользователя из контура без удаления самой учетной записи пользователя;
- удаление организационного контура с очисткой связанных наборов данных, документов, индексов, диалогов, агентов, коннекторов, MCP-серверов, токенов и настроек моделей.

2.2. Модуль выбора и подключения провайдеров AI-моделей

Это один из ключевых модулей Semantix. Модели являются заменяемым вычислительным компонентом, а Semantix обеспечивает их регистрацию, проверку, хранение настроек и использование в сценариях:

- получение каталога доступных провайдеров моделей и типов моделей;
- подключение провайдера с API-ключом, базовым URL и параметрами модели;
- поддержка разных типов моделей: chat/LLM, embedding, rerank, image-to-text, speech-to-text и text-to-speech;
- проверка доступности модели перед сохранением настроек: тест генерации, embedding, rerank, image-to-text или TTS в зависимости от типа;
- хранение настроек моделей в привязке к организационному контуру;
- просмотр подключенных моделей, типа модели, имени, использованных токенов, базового URL и максимального контекста;
- добавление пользовательских моделей через OpenAI-compatible API (VLLM, LocalAI, HuggingFace и другие совместимые среды исполнения ИИ)
- поддержка локальных провайдеров, включая Ollama, Xinfernce, LocalAI, LM-Studio, GPUStack и аналогичные варианты;
- удаление отдельных моделей или провайдеров из настроек организационного контура;
- использование разных моделей для разных задач: генерация ответа, построение эмбеддингов, переранжирование и обработка изображений.

2.3. Модуль наборов данных, баз знаний и обработки документов

Набор данных (DataSet) может создаваться вручную пользователем или автоматически при создании коннектора. Наборы данных являются основной сущностью для накопления, обработки и поиска корпоративных знаний:

- создание и настройка баз знаний с именем, описанием, языком, владельцем и правами доступа;
- привязка embedding-модели к базе знаний;
- загрузка документов и хранение сведений о типе, источнике, размере, местоположении, владельце и статусе обработки;
- настройка парсера: диапазон страниц, размер фрагмента, разделители, режим распознавания структуры, автоизвлечение ключевых слов и вопросов;
- разбиение документов на фрагменты, подсчет токенов и чанков;
- отслеживание прогресса обработки, сообщений об ошибках, длительности обработки и количества повторных попыток;
- хранение метаданных документа и использование их для фильтрации;
- хранение версий пайплайна обработки: версия парсера, версия чанкинга и использованная embedding-модель;
- настройка веса векторной схожести, порога схожести и веса даты для приоритизации более свежих документов;
- поддержка персональных и командных прав доступа к базе знаний.

2.4. Модуль поиска, RAG и проверяемых ответов

Модуль поиска и RAG обеспечивает получение релевантных фрагментов из наборов данных, формирование ответа на основе найденного контекста и возврат ссылок на первоисточники. Это принципиально отличает Semantix от простого чат-интерфейса к LLM – ответ формируется на основе управляемой базы знаний и может быть проверен пользователем:

- семантический поиск по одному или нескольким наборам данных;
- поиск по конкретным документам внутри набора данных;
- гибкая настройка top_k, top_n, порога схожести и веса векторной схожести;
- использование rerank-модели для уточнения порядка найденных фрагментов;
- поддержка ссылок на фрагменты и агрегированной информации по документам;
- поддержка режима с графом знаний при включении соответствующей настройки;
- дополнительные поисковые возможности: подсветка, keyword search, web search, related search, query mindmap и сводка при наличии соответствующих настроек;
- формирование ответов с источниками и ссылками на использованные фрагменты.

2.5. Модуль AI-ассистентов и диалоговых приложений

AI-ассистент в Semantix — это настраиваемое, диалоговое, приложение, которое использует выбранную модель, промпт, параметры генерации и одну или несколько баз знаний. Ассистенты предназначены для консультаций, внутреннего поиска, поддержки пользователей, анализа документов и предоставления ответов через интерфейс или API:

- создание ассистента с названием, описанием, иконкой и языком;
- выбор ИИ модели и rerank-модели;
- настройка параметров генерации
- задание системного промпта, приветствия, параметров и текста ответа при отсутствии релевантного контента;
- привязка ассистента к одному или нескольким наборам данных;
- привязка к саб-агентам конструктора БП
- настройка фильтрации по метаданным;
- ведение истории диалогов, сообщений, ссылок на источники и фактов;
- загрузка файлов непосредственно в диалог с последующим парсингом вложений;
- создание API-токенов для доступа к диалогам и интеграции с внешними приложениями;
- включение или отключение ссылок на источники в ответах.

2.6. Модуль AI-агентов и визуального конструктора

Конструктор бизнес-процессов – основной механизм построения агентных сценариев. Сценарий описывается как граф компонентов, связанных направленными переходами. Такой подход позволяет создавать не только одиночный чат-бот, но и управляемую цепочку обработки: поиск, анализ, генерация, ветвление, уточнение у пользователя, вызов инструментов и завершение сценария:

- хранение сценария: компоненты, параметры компонентов, связи, путь выполнения, история, память, результаты поиска и глобальные переменные;
- поддержка глобальных переменных сценария: запрос пользователя, пользователь, номер хода диалога, прикрепленные файлы;
- загрузка и проверка параметров компонентов перед выполнением;
- поддержка ветвлений через категоризацию и switch-логику;
- поддержка итерационных компонентов и шагов с дополнительным вводом пользователя;
- обработка ошибок компонентов с переходом на альтернативную ветку или возвратом значения по умолчанию;
- работа с файлами: изображения передаются как base64, документы обрабатываются перед передачей в сценарий;

- ведение ссылок на источники в виде фрагментов и агрегатов документов;
- ведение памяти и истории диалога в рамках сценария;
- трассировка вызовов инструментов: имя инструмента, аргументы, результат и время выполнения.

2.7. Мультиагентная логика и инструменты агента

Компонент Agent позволяет создавать специализированных агентов, которым передаются задача, обоснование выбора агента и контекст. Агент может использовать встроенные инструменты, подключенные компоненты и инструменты MCP-серверов:

- описание агента как функции с параметрами «Пользовательский промпт», «Режим рассуждения» и context;
- настройка максимального количества раундов рассуждения и вызова инструментов;
- привязка агента к модели организационного контура;
- подключение набора инструментов, реализованных как компоненты Semantix;
- подключение инструментов из MCP-серверов и преобразование их метаданных к формату вызова инструментов;
- потоковая генерация ответа при наличии downstream-компонента сообщения;
- ReAct-подобное выполнение: анализ запроса, вызов инструмента, добавление результата в контекст, продолжение генерации;
- генерация цитированного ответа на основе ссылок, накопленных в конструкторе БП.

2.8. Модуль коннекторов и синхронизации данных

Коннекторы предназначены для подключения внешних систем и загрузки данных в базы знаний Semantix. Они позволяют не ограничиваться ручной загрузкой файлов, а строить постоянно обновляемый слой корпоративных знаний.

Группа возможностей	Описание
Единый реестр типов	Коннекторы регистрируются через общий реестр обработчиков и управляются через унифицированный фасад создания, просмотра, обновления и удаления.
Создание набора данных	При создании коннектора может автоматически создаваться связанный набор данных, куда будут попадать загруженные данные.
Состояние и прогресс	Для коннектора хранятся статус, состояние, статистика, прогресс и ошибки инициализации/синхронизации.
Синхронизация	Поддерживаются первичная загрузка, backfill, режимы событийной или периодической синхронизации, обработка вложений и повторные попытки.

Группа возможностей	Описание
Очистка	При удалении коннектора может выполняться остановка задач, удаление секретов, индексов, загруженных сообщений и, при необходимости, связанного набора данных.

В Semantix предусмотрена регистрация следующих встроенных типов коннекторов, список которых расширяется:

- generic - универсальный коннектор с сохранением конфигурации и состояния;
- Zulip - загрузка сообщений, вложений, агрегация окон, формирование сводок, учет потоков и тем;
- Confluence - подключение корпоративной базы знаний;
- Google Drive, OneDrive, Yandex Disk, SMB, SharePoint - подключение файловых хранилищ;
- Outlook и IMAP - подключение почтовых источников;
- 1С – Коннектор к 1С системе
- Azure DevOps, Jira, GitHub – подключение систем разработки, задач и репозитория;
- Microsoft Teams – подключение коммуникационного источника;
- BDU FSTEC – специализированный источник данных для задач информационной безопасности;
- MTR – специализированный коннектор, реализованный в составе продукта.

2.9. Модуль MCP-серверов и внешних инструментов

MCP-модуль позволяет подключать внешние инструменты и сервисы по протоколу Model Context Protocol. Это расширяет возможности агентов: они могут не только отвечать на вопросы, но и обращаться к внешним инструментам, базам данных, корпоративным API и специализированным сервисам, а также управлять ими:

- создание MCP-сервера в рамках организационного контура;
- указание имени, URL, типа сервера, описания, переменных и HTTP-заголовков;
- проверка сервера и получение списка доступных инструментов перед сохранением;
- кэширование списка инструментов и признака enabled для каждого инструмента;
- импорт и экспорт набора MCP-серверов;
- просмотр списка серверов и деталей сервера;
- тестовый вызов инструмента с аргументами;
- использование инструментов MCP внутри Agent-компонента Canvas;
- поддержка авторизации через API key или Authorization header в зависимости от режима подключения.

2.10. Модуль API, SDK и внешней интеграции

Semantix предоставляет API и SDK-уровень для программного управления наборами данных, чатами, ассистентами, коннекторами и другими объектами. API-токены создаются в привязке к организационному контуру и конкретному источнику использования: агент, диалог или общий доступ:

- управление наборами данных и документами через API;
- запуск диалогов и получение ответов ассистентов;
- интеграция внешних приложений через API-токены;
- управление коннекторами и их конфигурацией;
- использование MCP-инструментов из внешних систем;
- поддержка сценариев встраивания Semantix в корпоративные приложения, порталы и внутренние сервисы.

ПЕРЕЧЕНЬ СОКРАЩЕНИЙ

AD – Active Directory, службы каталогов корпорации Microsoft для операционных систем семейства Windows Server.

AI – Artificial Intelligence, искусственный интеллект.

API – Application Programming Interface, интерфейс программирования приложений.

DB – Database, база данных.

DNS – Domain Name System, распределённая иерархическая система, которая преобразует человекочитаемые доменные имена в IP-адреса, необходимые для установления соединений в интернете.

GPU – Graphics Processing Unit, специализированный вычислительный чип, оптимизирован для параллельной обработки большого числа однотипных операций, что делает его эффективным для задач, требующих одновременного выполнения множества вычислений.

IAM – Identity and Access Management, управление идентификацией и доступом.

JSON – JavaScript Object Notation, текстовый формат для хранения и обмена структурированными данными.

LDAP – Lightweight Directory Access Protocol, протокол прикладного уровня, который служит «языком» для обмена данными между клиентскими приложениями и серверами, хранящими структурированную информацию (каталогами).

LLM – Large Language Model, большая языковая модель, продвинутый тип искусственного интеллекта, который специализируется на работе с человеческим языком: понимает его, анализирует и умеет генерировать связный, осмысленный текст.

MCP – Model Context Protocol, открытый стандарт с открытым исходным кодом, который стандартизирует обмен контекстом между большими языковыми моделями (LLM) и внешними источниками данных, инструментами и системами.

NTP – Network Time Protocol, сетевой протокол для синхронизации внутренних часов компьютеров, серверов, сетевого оборудования и других устройств в сети с использованием сетей с переменной задержкой (латентностью).

PDF – Portable Document Format, формат файлов, разработанный компанией Adobe в 1992 году для представления документов в едином виде, независимо от программного обеспечения, аппаратного обеспечения и операционной системы.

RAG – Retrieval-Augmented Generation, генерация с дополненной выборкой, технология, которая объединяет большую языковую модель (LLM) с внешними источниками информации для повышения точности и актуальности ответов.

SDK – Software Development Kit, набор инструментов, библиотек, документации и примеров кода, которые помогают разработчикам создавать приложения для определённой платформы, операционной системы или языка программирования.

TLS – Transport Layer Security, криптографический протокол, который обеспечивает безопасную передачу данных между клиентом и сервером в сети.