Руководство по работе с базой знаний (RAG)



Структура таблицы базы знаний

Представьте базу знаний как умный склад информации. Когда нужно полностью "перезагрузить" этот склад, мы поступаем просто:

1. **Полная очистка базы знаний**: удаляем старую таблицу и создаём точно такую же заново
   - Это как снести старый склад и построить новый по тому же чертежу
   - Гарантирует, что не останется "мусора" от предыдущих данных
2. **Размер вектора (embedding)**: может изменяться в зависимости от выбранного алгоритма векторизации
   - Вектор — это числовое представление смысла текста
   - Разные алгоритмы создают векторы разной длины (как разные форматы фотографий)

417	Да	Да	Да	gpt_knowledge_base	База знаний

Структура таблицы (JSON словарь):

АКТИВНО	СИХР. СТРУКТУРЫ	НАИМЕНОВАНИЕ	ЗАГОЛОВОК	ТИП	ОБЯЗАТЕЛЬНОЕ	ЗНАЧЕНИЕ ПО УМОЛЧАНИЮ	РАЗМЕР	Т
Да	Да	id	ID	AUTOINCREMENT	Да	NULL		
Да	Да	context	Домен	TEXT	Да			
Да	Да	content	Контент	TEXTAREA				
Да	Да	embeddings	embeddings	VECTOR		NULL	1536	

Создание компонента базы знаний

Важно: всегда создавайте компонент для базы знаний, на который будут ссылаться агенты.

Думайте о компоненте как о "визитной карточке" вашей базы знаний:

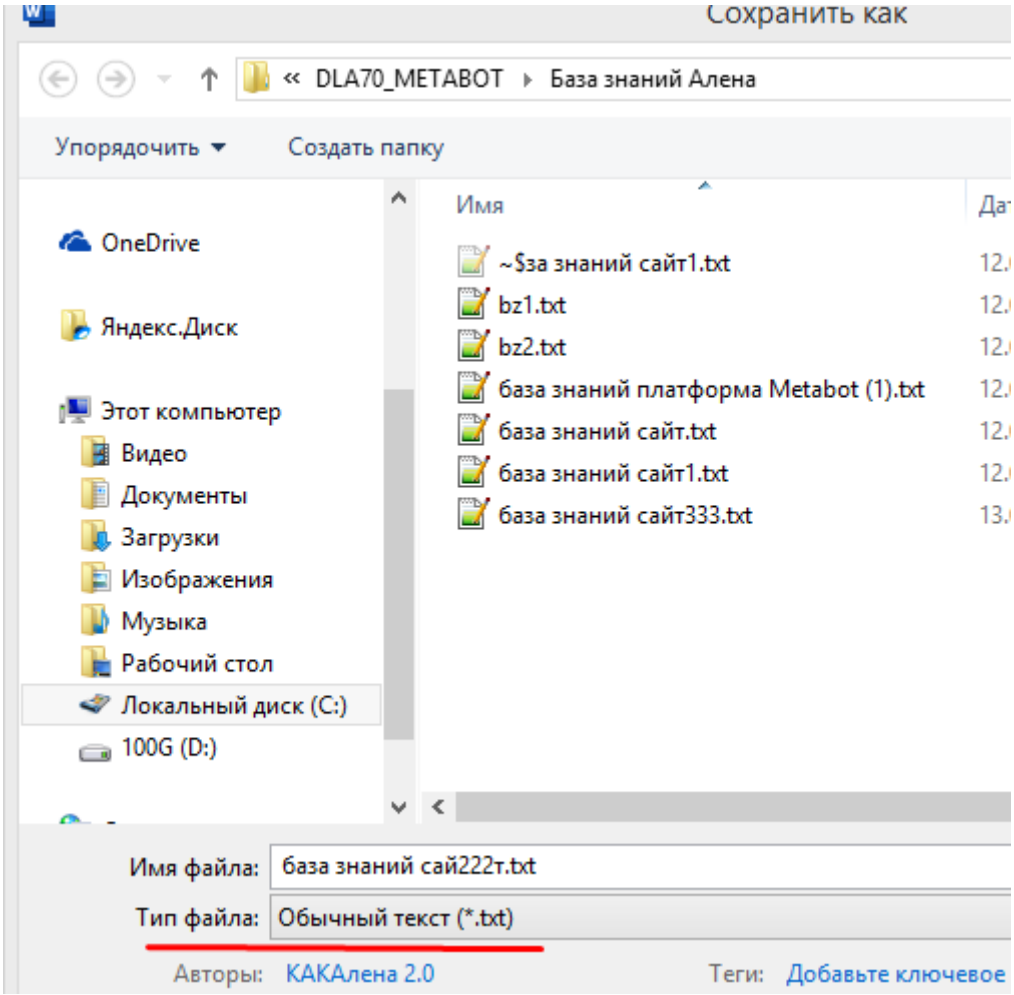
Скрипт	Тип	Свойства
AgentsLLM (5)		
MBQuery_run	Стандарт	Код: MBQuery_run Корневой скрипт: Да
menuKnowbases	Стандарт	Код: menuKnowbases Корневой скрипт: Да
vectorizeKnowbase	Стандарт	Код: vectorizeKnowbase Корневой скрипт: Да
createKnowbases	Стандарт	Код: createKnowbases Корневой скрипт: Да

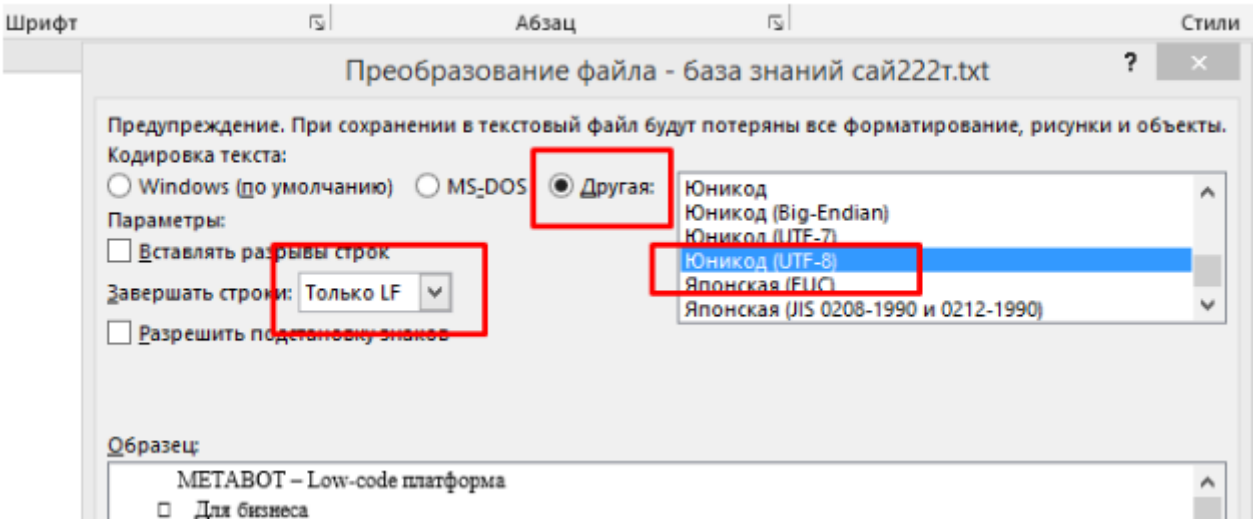
После создания в атрибутах появится компонент с описанием базы знаний:

	json	agent_MetabotServices_Attributes	{"ProviderName":"OpenAI","Params":{"model":"gpt-3.5-turbo","temperature":1}}	2(0:
	json	knowbase_defKnowBase	{"tableName":"gpt_knowledge_base","kbLimit":3,"kbMaxDistance":0.5,"providerName":"OpenAI_embeddings","params":{"model":"text-embedding-3-large","dimensions":1536}}	2(0:
07 - YURII	variable	activeAgent	MetabotInfo	2(

Подготовка текстового файла

Перед загрузкой файл нужно правильно "подписать" — добавить служебную информацию в начало:



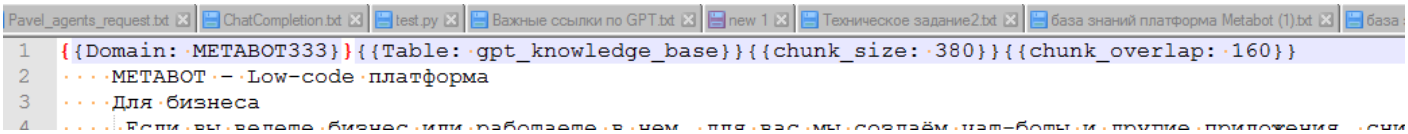


Обязательная служебная строка в начале файла:

```
{{Domain: METABOT33}}{{Table: gpt_knowledge_base}}{{chunk_size: 250}}{{chunk_overlap: 160}}
```

Эта строка работает как "инструкция по применению":

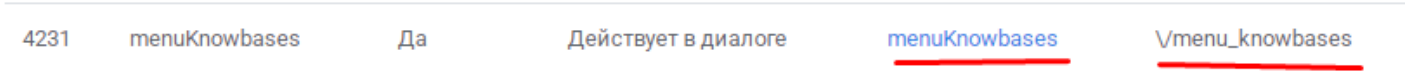
- `Domain` — область применения (как адрес склада)
- `Table` — название таблицы (как название полки на складе)
- `chunk_size` — размер кусочка текста для обработки (250 символов)
- `chunk_overlap` — перекрытие между кусочками (160 символов для связности)

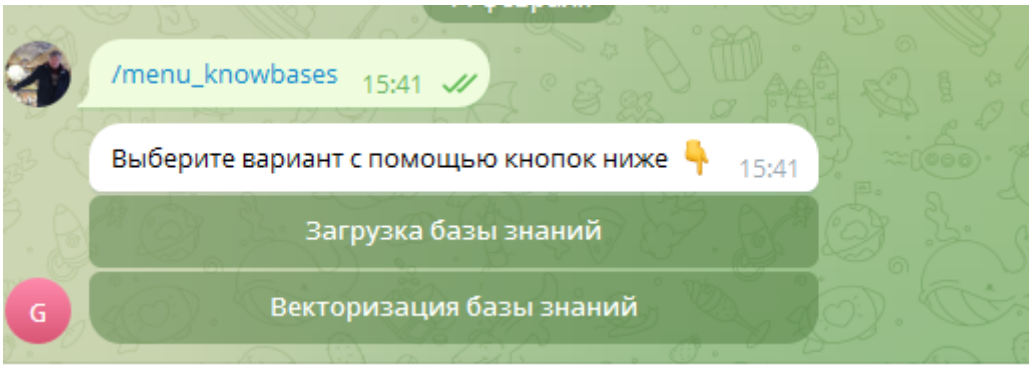


Загрузка и векторизация

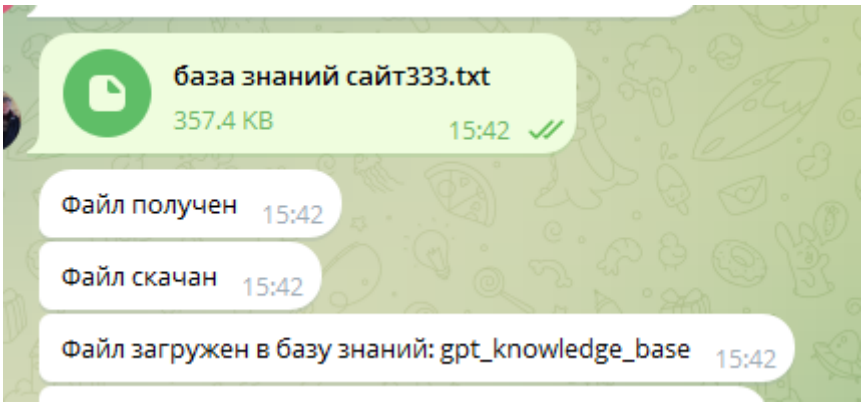
Процесс состоит из двух этапов:

Этап 1: Загрузка файла

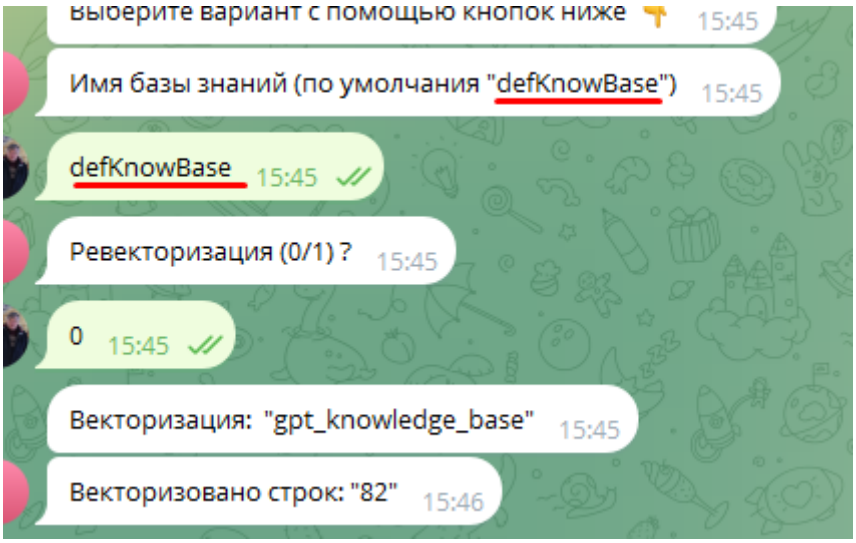




Система принимает подготовленный файл и сохраняет его содержимое.



Этап 2: Векторизация



Система превращает текст в числовые векторы для быстрого поиска по смыслу.

Важно: сохраните имя `defKnowBase` — оно понадобится в конфигурации в параметре `kbName`.

Совет: процесс загрузки и векторизации похож на работу переводчика — сначала он читает текст (загрузка), потом переводит его на "язык чисел" (векторизация), чтобы компьютер мог

быстро найти нужную информацию по смыслу, а не только по точным словам.

Версия #1
Павел Борисов создал 2025-09-05 14:07:07 UTC
Павел Борисов обновил 2025-11-12 14:03:47 UTC