

Описание программных модулей

Наименование ПО: NafanyaClassifierAI — Классификатор строительных ресурсов

Версия: 1.0

Дата документа: 02.07.2026

1. Общие сведения

Программный комплекс `res_cl` состоит из нескольких программных модулей — отдельных функциональных компонентов, взаимодействующих через REST API и общую СУБД.

2. Перечень модулей

2.1. Модуль `res_cl_api` (основной сервис)

Параметр	Значение
Тип	Серверное прикладное ПО
Язык	Python 3.12
Фреймворк	FastAPI
Порт	4300 (по умолчанию)
Назначение	Оркестрация классификации, веб-интерфейс, управление базой

Подмодули:

Подмодуль	Путь	Функция
API-классификация	<code>app/routers/api_classify.py</code>	Одиночная и пакетная классификация
API-база	<code>app/routers/api_base.py</code>	Загрузка, очистка, статус базы
API-ресурсы	<code>app/routers/api_resources.py</code>	Поиск по кодам
API-health	<code>app/routers/api_health.py</code>	Проверка работоспособности
UI	<code>app/routers/ui.py</code> , <code>app/templates/</code> , <code>app/static/</code>	Веб-интерфейс
Сервис классификации	<code>app/services/classify_service.py</code>	Бизнес-логика поиска

Подмодуль	Путь	Функция
Two-model index	app/services/two_models_index.py	In-memory кэш векторов
Импорт базы	app/services/ base_import_service.py	Фоновая загрузка и индексация
Предобработка	app/services/preprocessing.py	Нормализация текста group/nom
Клиент embed API	app/services/embedding_api.py	HTTP-клиент к embed-сервисам
Загрузчик файлов	app/services/file_loader.py	Парсинг xlsx/csv/json
БД	app/core/database.py, app/models/, app/repositories/	PostgreSQL + pgvector
Конфигурация	app/core/config.py	Настройки через pydantic-settings

2.2. Модуль `embed_res_nom_api`

Параметр	Значение
Тип	Сервис инференса ML-модели
Порт	3010
Модель	m33_nom_berta_best (BERT, 768 dim)
Назначение	Получение векторных представлений текстов для поиска номенклатур

API:

Метод	Путь	Описание
POST	/embed	Пакетное получение эмбедингов
GET	/health	Проверка работоспособности

Особенности:

- валидация входного текста (длина, детектор «мусорных» символов);
- возврат `problematic_indices` для некорректных строк;
- поддержка GPU (CUDA) и CPU.

2.3. Модуль `embed_res_group_api`

Параметр	Значение
Тип	Сервис инференса ML-модели

Параметр	Значение
Порт	3011
Модель	m34_group_berta_best (BERT, 768 dim)
Назначение	Получение векторных представлений текстов для поиска групп

API идентичен `embed_res_nom_api` (POST `/embed`, GET `/health`).

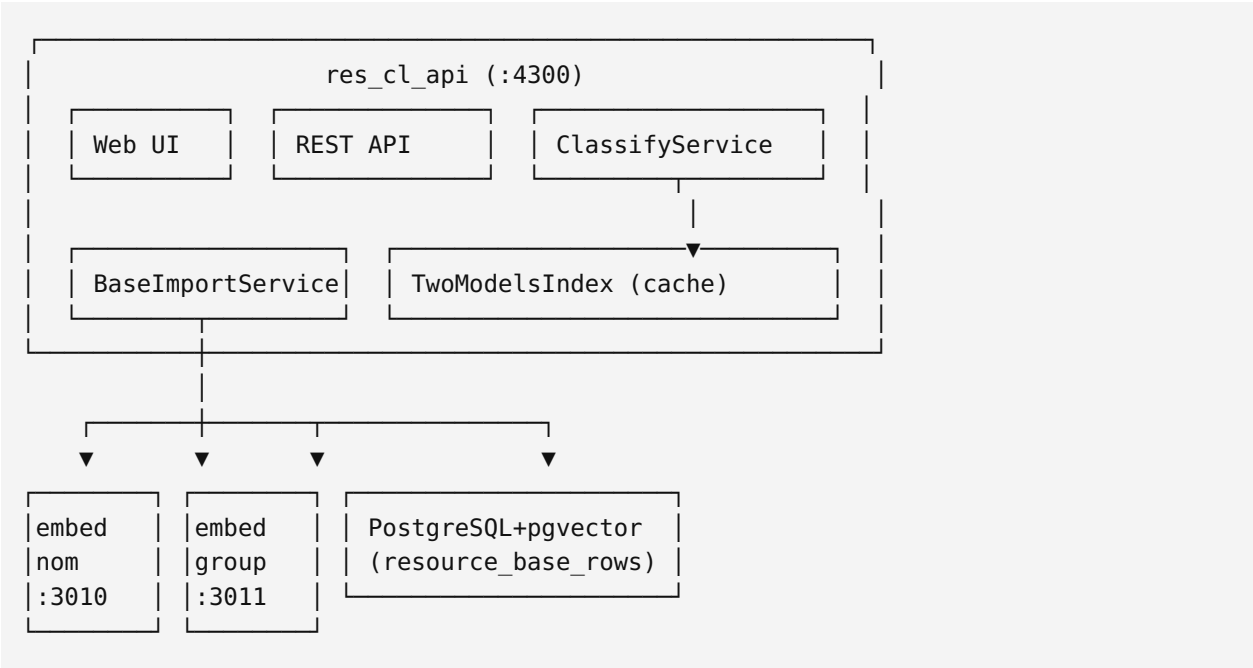
2.4. Модуль хранения данных (PostgreSQL + pgvector)

Параметр	Значение
Тип	СУБД с векторным расширением
Схема	<code>res_embed_api</code>
Основная таблица	<code>resource_base_rows</code>
Назначение	Хранение метаданных КСР, подготовленных текстов, эмбедингов

Индексы:

- B-tree: `group_code`, `nom_id`, `res_id`, `res_code`
- HNSW: `embedding_group`, `embedding_nom` (cosine)

3. Схема взаимодействия модулей



4. Потоки данных

4.1. Загрузка базы

1. `res_cl_api` читает файл (xlsx/csv/json).
2. Для каждого батча — предобработка текстов.
3. Запросы к `embed_res_nom_api` и `embed_res_group_api`.
4. Сохранение в PostgreSQL.
5. Построение in-memory кэша.

4.2. Классификация

1. Пользователь → `res_cl_api` (текст запроса).
2. Предобработка → `embed API` (query embeddings).
3. Поиск по кэшу → ранжирование → ответ.

5. Зависимости между модулями

Модуль	Зависит от
<code>res_cl_api</code>	<code>embed_res_nom_api</code> , <code>embed_res_group_api</code> , PostgreSQL
<code>embed_res_nom_api</code>	Файлы модели m33 на диске
<code>embed_res_group_api</code>	Файлы модели m34 на диске

`embed_res_*` модули работают автономно и могут использоваться другими приложениями через REST API.

6. Возможность автономного использования модулей

Модуль	Автономное использование
<code>res_cl_api</code>	Нет (требуется embed-сервисы и БД)
<code>embed_res_nom_api</code>	Да (REST API эмбедингов)
<code>embed_res_group_api</code>	Да (REST API эмбедингов)