Руководство по эксплуатации
ПО Платформа НейроСервис

Казань, 2025

1. Общие сведения

1.1. Назначение ПО Платформа НейроСервис

ПО Платформа НейроСервис – это набор алгоритмов, созданных для работы с Большими языковыми моделями (LLM), предназначенными для выполнения поисковых запросов по большому объёму текстовых данных (свыше 1 Гбайт) с учётом семантики языка.

1.2. Состав модулей ПО Платформа НейроСервис

ПО Платформа НейроСервис включает в себя следующие модули:

- Векторизация;
- Кластеризация;
- Классификация;
- Разбиение на чанки;
- Информационный поиск;

2. Подготовка к работе с ПО Платформа НейроСервис

2.1. Требования к рабочим станциям

2.1.1. Техническое обеспечение ПО Платформа НейроСервис

Для работы с ПО Платформа НейроСервис рабочие станции пользователей должны удовлетворять следующим минимальным требованиям к аппаратному обеспечению, приведенным ниже (Таблица 1).

Таблица 1 – Требования к конфигурации аппаратного обеспечения клиентской части

Компонент	Минимальная конфигурация
Вычислительная система	Nvidia Tesla T4 (16 Гб), 6 CPU по 2 ГГц
Оперативная память	16 Гб
Жесткий диск	6 Гб
Видеоадаптер	-
Сетевая плата	Ethernet 100 Мбит
Дополнительное оборудование	Монитор с разрешением не менее 1600x1200 пикселей, мышь, клавиатура

2.1.2. Программное обеспечение для подключения ПО Платформа НейроСервис

Для работы с ПО Платформа НейроСервис рабочие станции пользователей должны удовлетворять следующим минимальным требованиям к программному обеспечению, приведенным ниже (Таблица 2).

Таблица 2 – Требования к конфигурации ПО клиентской части:

Компонент	Конфигурация
Операционная система	Debian 10 и выше
Интегрированные среды разработки (IDE)	Atom, PyCharm, Spyder, Sublime Text

2.2. Подключение ПО Платформа НейроСервис

Использование Платформы производится стандартным вызовом нужного объекта через import или from.

import NeuroServis

from NeuroServis import VectorDB

3. Описание операций ПО Платформа НейроСервис

В начале необходимо произвести создание базы данных командой

VectorDB.create_database(path, model), указав в параметре path полный путь до директории содержащей файлы поддерживаемого формата (txt, pdf, docx, doc) и model название модели LLM.

Для обновления базы данных, необходимо выполнить команду VectorDB.update_database(path, model), указав в параметре path полный путь до директории содержащей файлы поддерживаемого формата (txt, pdf, docx, doc), которые необходимо дополнительно проиндексировать, и model название модели LLM.

После создания базы данных будет доступна возможность произвести следующие операции:

- Кластеризация. Clustering.create_clusters(min_cluster_size, min_samples), где
    - min_cluster_size - минимальное количество точек, необходимое для формирования кластера. Этот параметр контролирует минимальный размер кластера;
    - min_samples - количество соседей, которое используется для оценки локальной плотности точки. Чем выше значение, тем более "плотными" должны быть области, чтобы считаться кластерами.
- Классификация. Classify.distribute(classes),
  где
    - classes - список существующих классов.

- Семантический поиск. Search(type, query),
  где
  - type - тип поиска. semantic - смысловой поиск, fulltext - полнотекстовый поиск, lps - гибридный метод поиска с выделением внимания из слоев языковой модели.

Перечень сокращений

CPU (англ. Central Processing Unit) – центральное обрабатывающее устройство

LLM (с англ. Large Language Model) – большая языковая модель

ПО – программное обеспечение