Министерство образования Республики Беларусь

Учреждение образования

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

Факультет Информационных технологий и управления

Кафедра Интеллектуальных информационных технологий

**ОТЧЁТ**

по дисциплине «Естественно-языковой интерфейс интеллектуальных систем»

Лабораторная работа №6

Тема: Диалоговая система с поддержкой естественного языка

Выполнили: Демидовец Д.В.

Козырев Д.А.

гр. 221703

Проверил: Крапивин Ю.Б.

Минск 2025

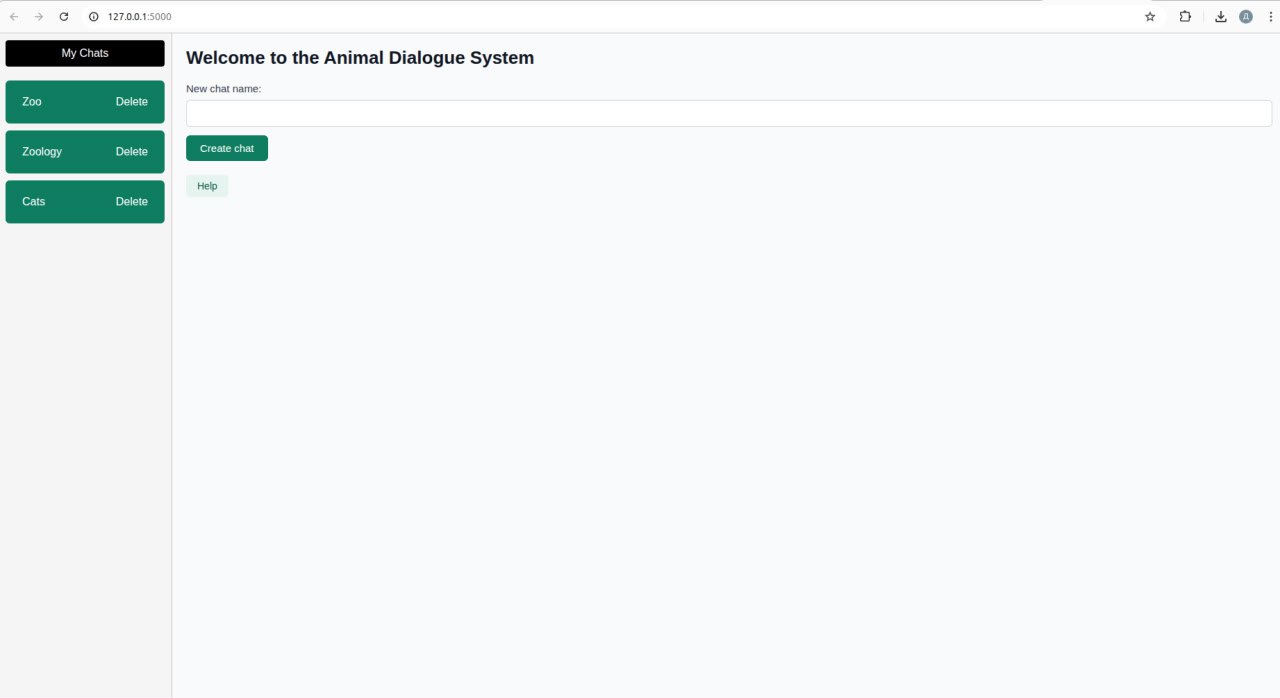
**Цель работы:** освоить принципы разработки диалоговых систем с поддержкой естественного языка.

**Задачи лабораторной работы: ­**

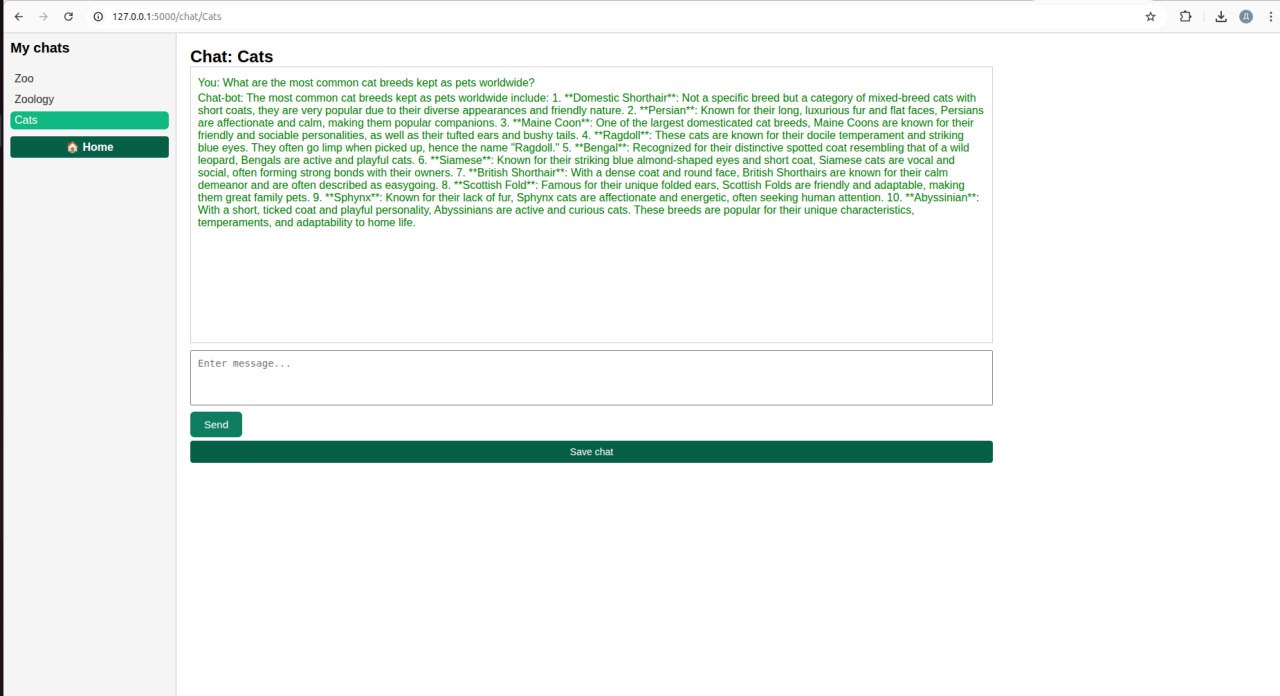
­ изучить основы создания диалоговых систем с поддержкой естественного языка;

­ закрепить навыки программирования при решении задач организации диалогового взаимодействия с поддержкой естественного языка.

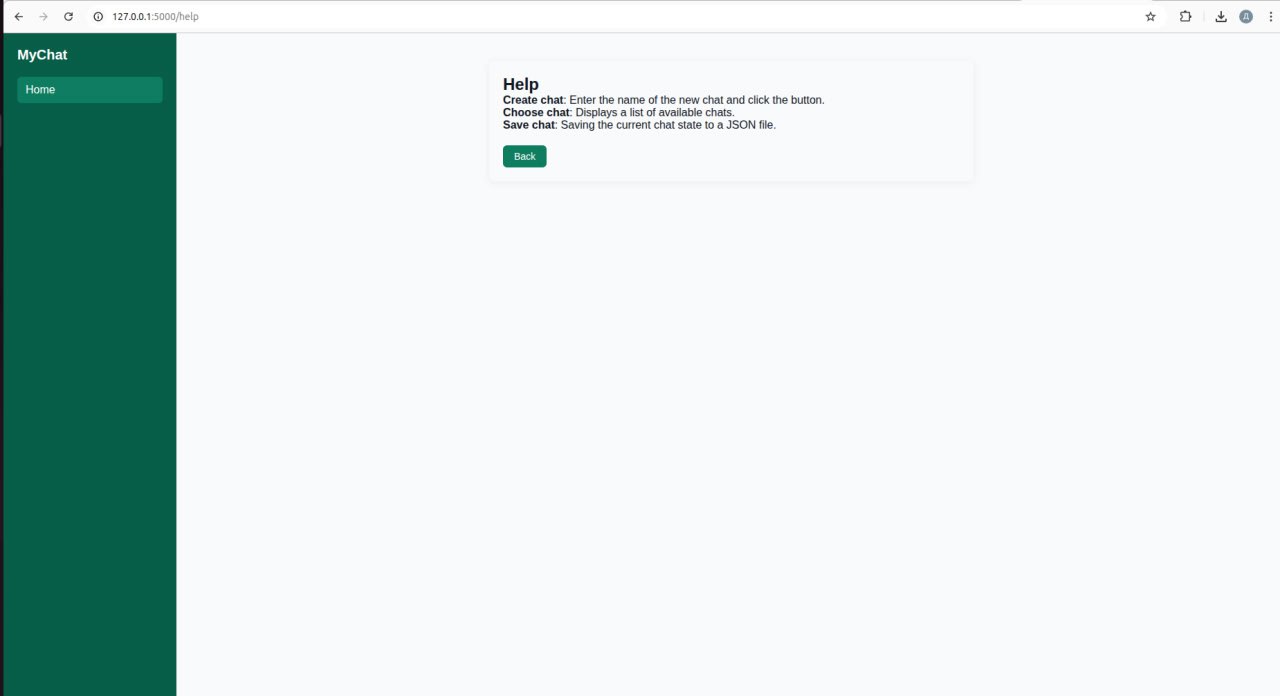
**Предметная область:** Животные, язык: английский



Главный экран

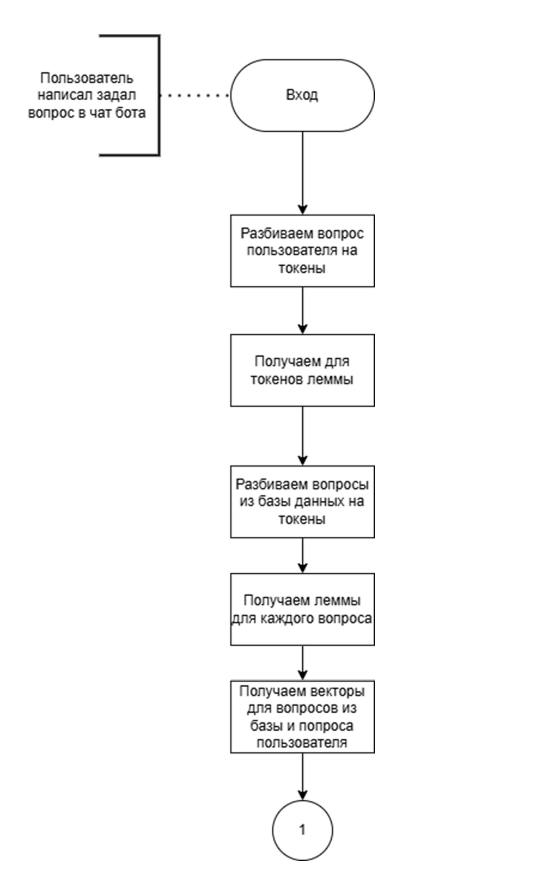


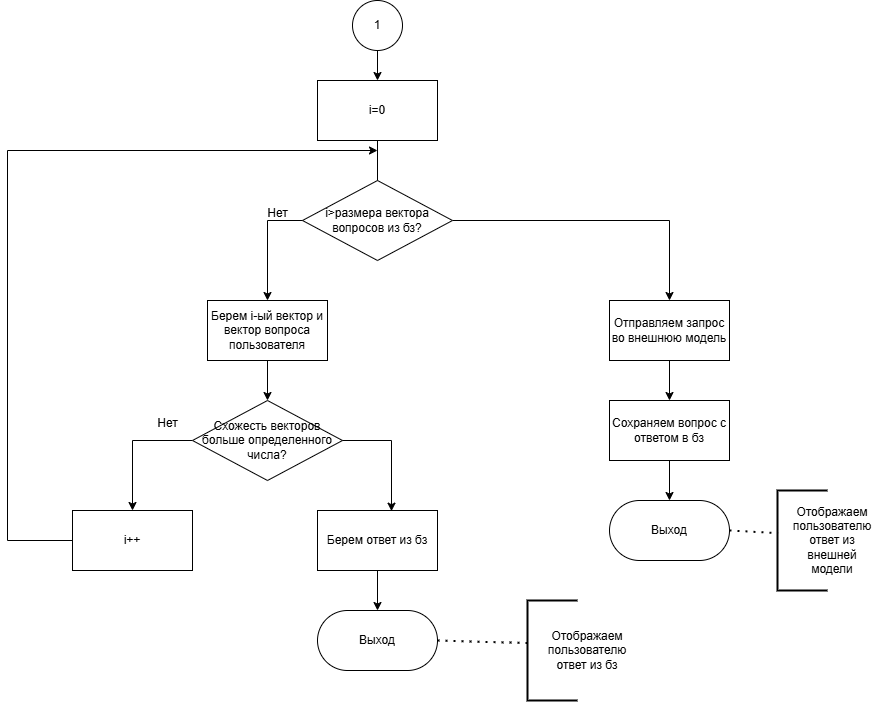
Экран с чатом



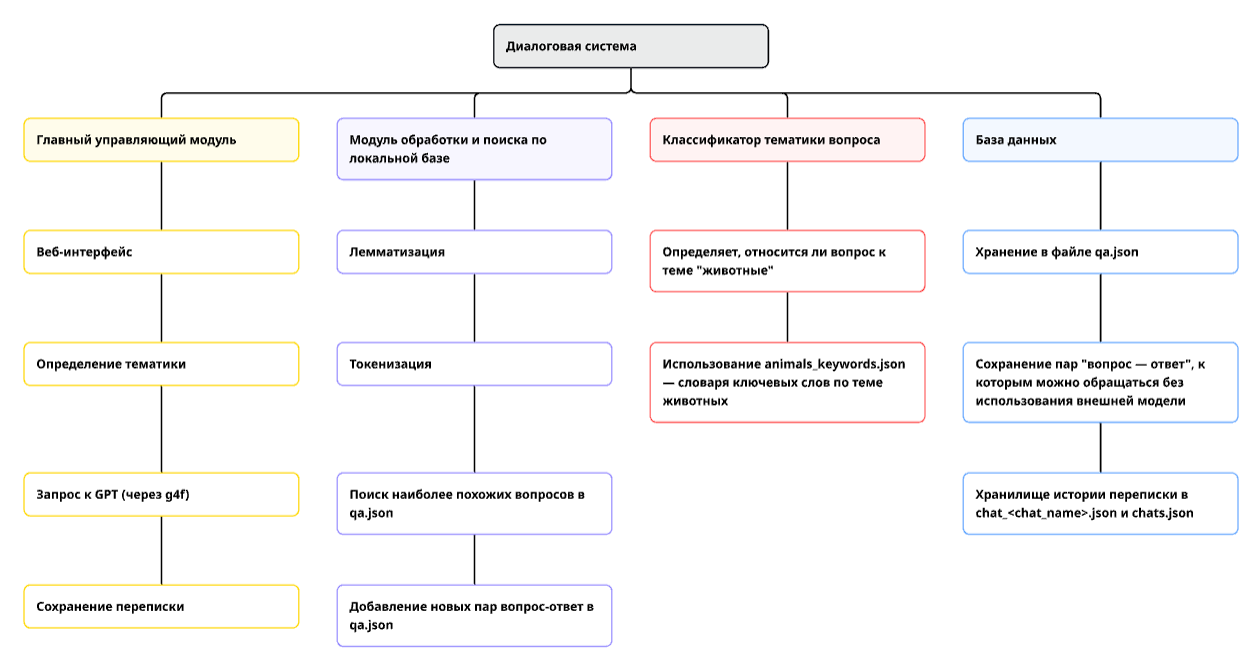
Экран с помощью

**Логическая структура сценария диалога:**

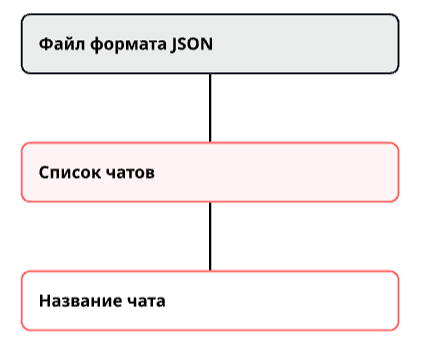
****



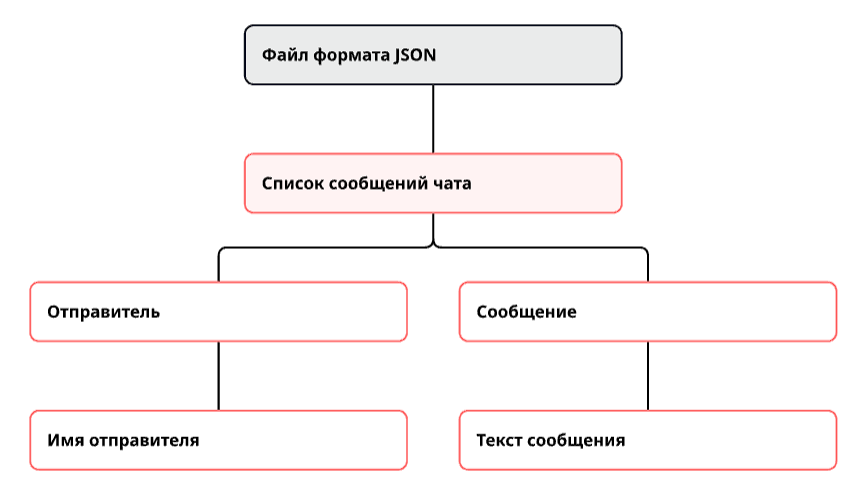
**Структурно-функциональная схема приложения:**

****

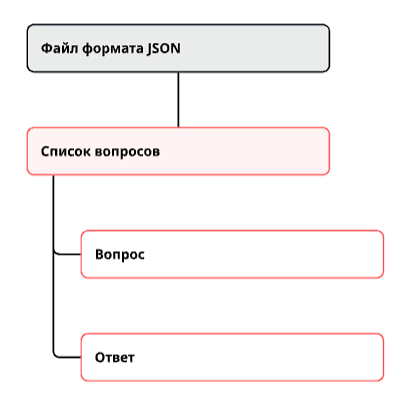
**Структуры хранения данных:**

****

Структура хранения названий чатов



Структура хранения данных чата

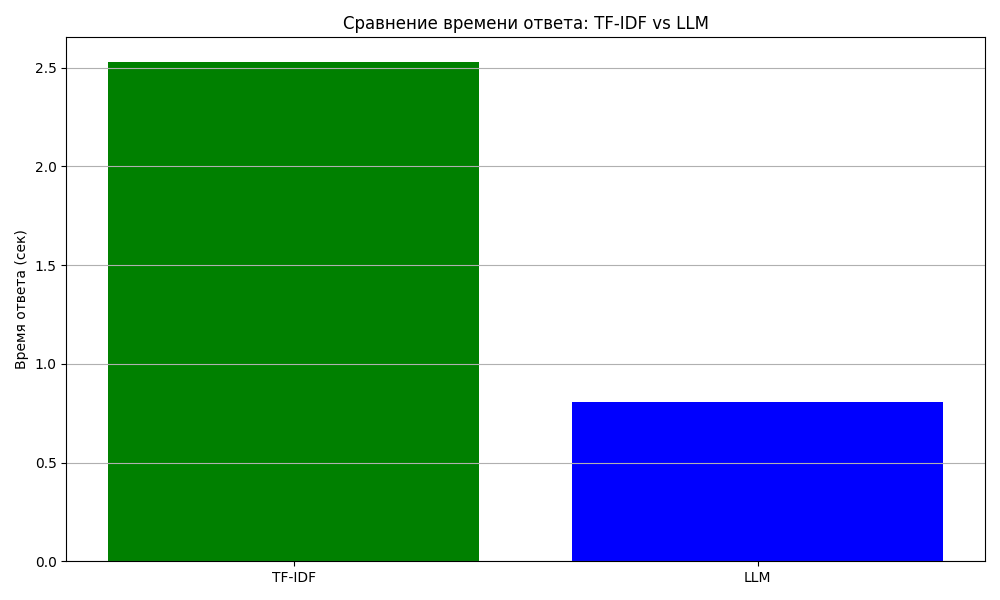


Структура базы данных с ответами на вопросы

**Список используемых библиотек:**

* Flask – для создания веб-приложения и API.
* os – для работы с файловой системой (создание, проверка, удаление файлов).
* json – для чтения и записи данных в формате JSON.
* nltk – для обработки естественного языка:
  + word\_tokenize – токенизация текста,
  + WordNetLemmatizer – лемматизация,
  + stopwords – стоп-слова,
  + punkt и wordnet – загрузка ресурсов.
* string – для работы с пунктуацией (фильтрация символов).
* sklearn (scikit-learn) – для обработки текстов и оценки схожести:
  + TfidfVectorizer – векторизация текста,
  + cosine\_similarity – расчет косинусного сходства.
* g4f – для взаимодействия с языковой моделью GPT (альтернатива OpenAI API).
  + g4f.client – клиент для генерации ответов от модели.

**Оценка быстродействия приложения**



Результаты сравнения показали, что время ответа при использовании TF-IDF-механизма оказалось **дольше**: в среднем около **2.5 секунд**, тогда как при обращении к внешней языковой модели (LLM), даже с учетом имитации сетевой задержки, ответ формировался **менее чем за 1 секунду**. Это может быть связано с тем, что при TF-IDF каждый раз происходит полная предобработка всех сохранённых вопросов и повторная векторизация, что создаёт избыточную нагрузку. В то время как вызов языковой модели, особенно если он хорошо оптимизирован или кэшируется, может давать ответы быстрее. Таким образом, несмотря на распространённое мнение о медлительности LLM, в данной реализации **она оказалась быстрее** локального поиска по базе.

**Вывод:**

В ходе выполнения лабораторной работы было разработано веб-приложение на основе фреймворка Flask, реализующее диалоговую систему, способную отвечать на вопросы пользователя на тему животных. Основной функционал системы включает: определение релевантности вопроса теме животных, поиск ответа в локальной базе ранее заданных вопросов с использованием TF-IDF и косинусного сходства, а при отсутствии ответа — обращение к языковой модели GPT для генерации ответа и последующее сохранение нового вопроса и ответа в базу. Реализована структура хранения чатов и сообщений с возможностью создания, удаления и ведения истории диалога. Разработанная система демонстрирует интеграцию средств предварительной обработки текста, алгоритмов машинного обучения и возможностей современных языковых моделей. Полученные в процессе работы навыки подтверждают возможность создания гибких и расширяемых интеллектуальных систем, работающих с естественным языком.