

Секция «Цифровая трансформация образования и новые технологии обучения»

Разработка многоязычного чат-бота на базе Phi3 и Yandex Translate API

Научный руководитель – Конкин Никита Александрович

Ермохин Андрей Алексеевич

Студент (бакалавр)

Поволжский государственный технологический университет, Радиотехнический
факультет, Йошкар-Ола, Россия

E-mail: imyourinnerself@yandex.ru

В настоящее время возрастает потребность в доступных и локализованных образовательных инструментах для иностранных студентов, сталкивающихся с языковым барьером и ограниченным доступом к платным сервисам. Локально работающий чат-бот с многоязычной поддержкой, включая языки народов России, решает проблему стоимости и зависимости от облачных решений, а интеграция с Yandex Translate обеспечивает точную коммуникацию.

Цель работы — разработка многоязычного чат-бота для взаимодействия с иностранными студентами. Задачи — разработка графического интерфейса, реализация перевода сообщений на язык пользователя и сохранение истории сообщений.

Техническая реализация проекта включает фреймворк Angular для клиентской части, фреймворк Flask для серверной части и языковую модель Phi3-4k-mini-onnx в роли чат-бота. [1, 2, 3]

В результате разработки был успешно реализован интерфейс для общения с чат-ботом Phi3. При отправке запросов пользовательское сообщение отображается на экране вместе с датой отправки. Серверная часть сохраняет сообщение в историю (JSON-файл) и переводит его на английский язык через API Yandex Cloud Translate. Затем к переведённому сообщению добавляются инструкции, и оно отправляется на сервер Phi3 для генерации ответа. Полученный ответ переводится на язык пользователя, сохраняется в историю сообщений и отображается на экране. Благодаря этому иностранные студенты могут взаимодействовать с чат-ботом на своём родном языке, используя его как помощника в учебном процессе.

Разработанный чат-бот имеет два ключевых преимущества: во-первых, он работает локально на компьютере пользователя, что делает его основной функционал полностью бесплатным; во-вторых, он использует сервис Yandex Translate, обеспечивающий точный перевод на более чем 100 языков, включая языки народов России. Благодаря этому сочетанию локальной работы и доступного перевода, решение получается значительно экономичнее существующих аналогов.

Дальнейшим развитием проекта является дообучение (fine-tuning) модели для работы с пользовательскими статьями, в частности с лекционными материалами, написанными преподавателями учебного заведения.

Источники и литература

- 1) Гек Д.К. Angular как средство создания нативного приложения // Актуальные проблемы авиации и космонавтики. Сборник материалов VII Международной научно-практической конференции, посвященной Дню космонавтики. 12–16 апреля 2021 г. Красноярск, 2021. Т. 2. С. 486–487.
- 2) Косых Н.Е., Хомоненко А.Д., Куранова О.Н. Разработка Web-приложения для анализа настроений текста с помощью фреймворка Flask и языка Python // Наукоемкие технологии в космических исследованиях Земли. 2022. Т. 14. № 1. С. 45-52

3) huggingface.co: <https://huggingface.co/microsoft/Phi-3-mini-4k-instruct-onnx>