

УДК 004.9

РАЗРАБОТКА ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ РАСПОЗНАВАНИЯ КЛЮЧЕВЫХ СЛОВ В ТЕКСТЕ НА ОСНОВЕ АНАЛИЗА УСТНОЙ РЕЧИ

Берёзкин Д.А. (ТИУ), Анпилогов А.Б. (ТИУ)

Научный руководитель – старший преподаватель Михайлова О.В. (ТИУ)

Введение. В настоящее время существует потребность внедрения информационной системы, направленную на оптимизацию бизнес-процесса подачи заявок специалистам. Современные системы планирования ресурсов (ERP) предоставляют недостаточно эффективные решения. Основной задачей является улучшение процесса формирования и подачи заявок путём создания информационной системы обработки голосового сообщения заявителя, вычленения из него ключевых слов, и, ориентируясь по ним, направление автоматически созданной заявки нужному специалисту.

Основная часть. Была разработана информационная система, которая выполняет следующие задачи:

- 1) Преобразование речи в текст с использованием претренированной модели нейронной сети Vosk;
- 2) Разбиение текста, по ключевым словам, согласно предметной области технической поддержки организации;
- 3) Создание системой записи в базу данных, хранящуюся на сервере, по выделенным ключевым словам.

Для расширения системы в дальнейшем, была создана подсистема формирования собственного датасета – алгоритм сохраняет голосовое сообщение и результат обработки нейронной сетью.

В работе используется модель Vosk, представляющая собой библиотеку для распознавания речи. Модели Vosk используют рекуррентные нейронные сети (RNN) и другие архитектуры, такие как трансформеры, для анализа последовательностей аудиоданных и извлечения признаков.

Модели Vosk могут включать следующие типы слоев:

- 1) Входной слой принимает обработанные аудиосигналы в виде спектрограмм;
- 2) Скрытые слои составляют рекуррентные слои (RNN), слои LSTM (Long Short-Term Memory). RNN используются, когда важно соблюдать последовательность, когда важен порядок поступающих объектов, в частности в обработке аудио информации [2]. LSTM-сеть хорошо приспособлена к обучению на задачах классификации, обработки и прогнозирования временных рядов в случаях, когда важные события разделены временными задержками с неопределённой продолжительностью и границами [1]. Слои Attention позволяют модели фокусироваться на определённых частях выходных данных;
- 3) Выходной слой формирует вероятностное распределение по всем возможным символам или словам на основе выходов скрытых слоёв.

Модель Vosk-model-ru-0.22, используемая в работе, имеет следующие показатели по метрике Word Error Speed (WER) / Speed: 13.35 (аудиокниги open_stt); 20.73 (open_stt youtube); 37.38 (звонки openstt); 19.71 (устройства sova) [3].

Выводы. Проведён анализ существующих решений, разработана собственная информационная система для обработки голосовых заявок пользователей и направлению заявки нужному специалисту.

Список использованных источников:

1. Долгая краткосрочная память / [Электронный ресурс] // Википедия : [сайт]. — URL:

https://ru.wikipedia.org/wiki/Долгая_краткосрочная_память (дата обращения: 08.02.2025).

2. Рекуррентные нейронные сети / [Электронный ресурс] // ifmo : [сайт]. — URL: https://neerc.ifmo.ru/wiki/index.php?title=Рекуррентные_нейронные_сети#.D0.9E.D0.B1.D0.BB.D0.B0.D1.81.D1.82.D0.B8_.D0.B8_.D0.BF.D1.80.D0.B8.D0.BC.D0.B5.D1.80.D1.8B_.D0.BF.D1.80.D0.B8.D0.BC.D0.B5.D0.BD.D0.B5.D0.BD.D0.B8.D1.8F (дата обращения: 01.02.2025).

3. Vask models / [Электронный ресурс] // Alphacephei : [сайт]. — URL: <https://alphacephei.com/vosk/models> (дата обращения: 15.01.2025).