

ОТЗЫВ НАУЧНОГО РУКОВОДИТЕЛЯ

на выпускную квалификационную работу обучающегося
Санкт-Петербургского государственного университета
Лишик Елены Викторовны
«Распознавание речи с использованием алгоритмов глубокого обучения»

Выпускная квалификационная работа Лишик Елены Викторовны посвящена созданию системы автоматического распознавания речи и конвертации ее в текст с использованием алгоритмов глубокого обучения.

Зачастую при общении в мессенджерах пользователи используют голосовые сообщения, однако получатель сообщения не всегда имеет возможность их прослушать. Перед Лишик Е.В. стояла задача разработать решение для осуществления автоматической конвертации аудиофайлов в текстовый вид, при этом должна была быть обеспечена интеграция такого сервиса в один из популярных на сегодняшний день мессенджеров. Данная задача является сложной и несомненно актуальной. Несмотря на достаточно большое количество готовых коммерческих решений в области распознавания речи, предлагаемых крупными компаниями такими как Яндекс, Google и др. и открытое API, их интеграция с современными мессенджерами невозможна.

Таким образом, для решения поставленной задачи Еленой Викторовной был разработан собственный сервис, состоящий из трех ключевых подсистем: сервера, выполняющего преобразование аудиодорожки в необходимый формат и передачу файла следующей подсистеме, непосредственно отвечающей за распознавание речи. Данный модуль в своей работе использует библиотеку распознавания речи с открытым исходным кодом (Sphinx), обладающую рядом преимуществ (поддержку русского языка, высокую скорость распознавания, а также наличие библиотек для Java). Третья подсистема является связующим элементом между пользовательским интерфейсом (Telegram) и сервисом преобразования «аудио-текст».

Работа Лишик Е.В. состоит из 4 глав. Первая глава содержит краткое описание истории развития технологии распознавания речи. Вторая глава посвящена описанию принципов работы серверного приложения и его реализации (подсистемы (1)). В третьей главе проведен сравнительный анализ основных систем распознавания речи, и как следствие сделан выбор в пользу сервиса Sphinx: описан его математический аппарат и поэтапно изложены шаги интеграции данного сервиса в разработанное приложение (подсистема (2)). В последней главе описано создание бота в качестве связующего элемента между пользователем и сервисом распознавания (подсистемы (3)).

Елена Викторовна провела достаточно широкий обзор литературы по рассматриваемой проблеме, как российских, так и зарубежных авторов, с учетом которых были выбраны алгоритмы и методы, лежащие в основе каждого из этапов работы системы распознавания.

С помощью разработанного программного средства удалось достичь достаточно хороших показателей точности распознавания отдельно взятых слов, однако данный показатель ниже для слитной речи, так как отсутствует необходимая по величине база для обучения. Тем не менее, считаю, что поставленная задача была успешно выполнена.

Выпускная работа Лишик Е.В. полностью удовлетворяет требованиям, предъявляемым к магистерским диссертациям, а ее автор заслуживает оценки «отлично».

Кандидат физ.-мат. наук, доцент кафедры
теории систем управления
электрофизической аппаратурой СПбГУ



А. Г. Головкина