

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
МЕХАНИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ
Кафедра дифференциальных уравнений и системного анализа

Аннотация к дипломной работе
РАСПОЗНАВАНИЕ УСТНОЙ РЕЧИ

Корнилова Дарья Витальевна

Научный руководитель:
кандидат физ.-мат. наук,
доцент Малевич А. Э.

2020

Дипломная работа содержит: 66 страницы, 11 рисунков, 2 таблицы, 17 источников, 3 приложения.

Ключевые слова: РАСПОЗНАВАНИЕ РЕЧИ, ASR, STT, SPEECH RECOGNITION, РАСПОЗНАВАНИЕ УСТНОЙ РЕЧИ, СРАВНЕНИЕ СИСТЕМ ПО РАСПОЗНАВАНИЮ РЕЧИ, ИНСТРУМЕНТЫ ДЛЯ ШУМОПОДАВЛЕНИЯ

Объектом исследования дипломной работы является системы по распознаванию устной речи.

Целями дипломной работы были поиск и сравнения систем для распознавания речи, улучшения их качества и реализация веб-сервера, который обеспечит взаимодействие с инструментом по распознаванию речи.

Для достижения поставленной цели был использован язык программирования Python. Три инструмента Google Speech API, Kaldi, ESPnet сравнивались относительно метрик WER, SER и SF, а также WER и SER для стеммированных слов. С помощью дополнительной предобработки аудиозаписей (удаление шума) удалось улучшить показатели WER и SER. Итогом работы стал RESTful сервер реализованный с помощью библиотеки Flask, а многопоточный режим с помощью `gevent.pywsgi.WSGIServer`.

В дипломной работе были получены следующие результаты:

1. При сравнении различных инструментов для распознавания речи, используя метрики, ESPnet показал себя лучше по SER и WER, а Kaldi по SF.
2. Сделав дополнительную предобработку аудиофайлов, было в некоторой степени улучшено качество распознавания.
3. Реализован REST-сервер на Python, который позволяет взаимодействовать с ESPnet для распознавания речи.

Дипломная работа является завершенной, поставленные задачи решены в полной мере, присутствует возможность дальнейшего развития исследований.

Дипломная работа выполнена автором самостоятельно.

The thesis contains 66 pages, 11 figures, 2 tables, 17 sources, 3 appendices.

Keywords: SPEECH RECOGNITION, ASR, STT, COMPARISON OF SPEECH RECOGNITION SYSTEMS, NOISE REDUCTION TOOLS.

The object of research of the thesis is systems for speech recognition.

The goals of the thesis were to search for and compare systems for speech recognition, improve their quality, and implement a web server that will provide interaction with the speech recognition tool.

To achieve this goal, the Python programming language was used. The three Google Speech API tools, Kaldi, and ESPnet were compared against the WER, SER, and SF metrics, as well as WER and SER for stemmed words. Using additional preprocessing of audio recordings (noise removal) I was able to improve the WER and SER indicators. The result was a RESTful server implemented using the Flask library, and multithreaded mode using `gevent.pywsgi.WSGIServer`.

The following results were obtained in the thesis:

1. When comparing different speech recognition tools using metrics, ESPnet performed better on SER and WER, and Kaldi on SF.
2. By making additional preprocessing of audio files, the recognition quality was improved to some extent.
3. Implemented a REST server in python that allows you to interact with ESPnet for speech recognition.

The thesis is completed, the tasks have been fully solved, and there is an opportunity for further development of research.

The thesis was completed by the author independently.