

TEAM PROJECT INFORMATION FOLDER – JUNE 2023

KATEDRA SYSTEMÓW MULTIMEDIALNYCH

Zespół projektowy: 7@KSMM'2023	1. Wiktor Krasieński - leader 2. Przemysław Rośień 3. Jakub Nowak 4. Jan Stopiński
Opiekun:	Prof. Andrzej Czyżewski
Klient:	Mgr inż. Szymon Zaporowski
Data zakończenia:	
Słowa kluczowe:	rozpoznawanie mowy; uczenie maszynowe; cyfrowe przetwarzanie sygnałów



TEMAT PROJEKTU:

Badanie algorytmów uczenia maszynowego w zastosowaniu do rozpoznawania mowy personelu medycznego

CELE I ZAKRES PROJEKTU:

Celem projektu jest opracowanie metodyki dotrenowania i testowania dostępnych algorytmów transkrypcji mowy na tekst, np. w środowisku do uczenia głębokiego DeepSpeech. Następnie, w oparciu o zadeklarowaną współpracę ze strony lekarzy Gdańskiego Uniwersytetu Medycznego należy zbudować i nagrać słownik polskich wyrażen medycznych, które są używane przy opisywaniu chorób, kierowaniu na badania przez specjalistów, wypisaniu recept. Nagrywanie należy przeprowadzić zarówno w warunkach korzystnych od strony akustycznej, jak również w hałaśliwym otoczeniu. Następnie, w ten sposób otrzymany materiał należy adnotować i użyć do dotrenowania dostępnej sieci neuronowej, która wcześniej już została wytrenowana w oparciu o słownik języka polskiego np. Mozilla Polish. W toku eksperymentów należy uzyskać wyniki oceny skuteczności rozpoznawania wyrazów związanych ze słownikiem medycznym. Korzystnie również w zróżnicowanych warunkach akustycznych. Dodatkowym elementem pracy może być utworzenie adaptacyjnych formularzy, umożliwiających ich wypełnianie głosowe przez lekarzy.

OSIĄGNIĘTE REZULTATY:

1. Systematyczny przegląd literatury.
2. Zapoznanie się z narzędziami speech-to-text.
3. Przegląd dostępnych narzędzi do adnotacji.
4. Wygenerowanie próbek słownictwa medycznego.
5. Testowanie dostępnych narzędzi speech-to-text na nagranych próbkach.
6. Zapoznanie się z metrykami transkrypcji mowy do tekstu.
7. Analiza uzyskanych wyników transkrypcji mowy medycznej dzięki wybranym narzędziom.
8. Wytworzenie i zgłoszenie artykułu na konferencje.

CECHY CHARAKTERYSTYCZNE ROZWIĄZANIA, KIERUNKI DALSZYCH PRAC:

Cechy charakterystyczne:

1. Wybrane narzędzia do tworzenia speech-to-text
 - Whisper (small, medium, large)
 - Google
 - Microsoft Azure
2. Wygenerowanie 1200 próbek słownictwa medycznego (100 słów) w rozszerzeniu mp3
 - 600 męskich
 - 600 żeńskich
 - W tym 100 męskich syntetycznych dla 600 męskich i 100 żeńskich syntetycznych dla 600 żeńskich
3. Wykorzystane metryki w analizie wyników

TEAM PROJECT INFORMATION FOLDER – JUNE 2023

- WIL
- WER
- CER
- MER- Levenshtein distance
- Jaccard distance Further

Kierunki dalszych prac:

1. Dogenerowanie większej ilości próbek słownictwa medycznego.
- Levenshtein distance
 - Jaccard distance Further

Kierunki dalszych prac:

1. Dogenerowanie większej ilości próbek słownictwa medycznego.
2. Dotrenowanie wybranego modelu.
3. Testowanie dotrenowanego modelu
4. Wykonanie Raportu końcowego.

TEAM PROJECT INFORMATION FOLDER – JUNE 2023

DEPARTMENT FULL NAME

Project team: 7@KSMM'2023	1. Wiktor Krasieński - leader 2. Przemysław Rośleń 3. Jakub Nowak 4. Jan Stopiński
Supervisor:	Prof. Andrzej Czyżewski
Client:	M.Sc.Eng. Szymon Zaporowski
Date:	
Key words:	speech recognition; machine learning; digital signal processing



PROJECT TITLE:

Investigation of machine learning algorithms for speech recognition in the context of medical personnel

OBJECTIVES AND SCOPE:

The aim of the project is to develop a methodology for retraining and testing available speech-to-text transcription algorithms, for example, in the DeepSpeech deep learning environment. Then, based on the declared cooperation with doctors from the Medical University of Gdańsk, it is necessary to build and record a dictionary of Polish medical expressions used in describing diseases, referring patients for tests by specialists, and issuing prescriptions. The recording should be conducted in both favorable acoustic conditions and noisy environments. Subsequently, the obtained material should be annotated and used to retrain an existing neural network that has previously been trained based on a Polish language dictionary, such as Mozilla Polish. The experiments should yield results evaluating the effectiveness of recognizing words related to the medical dictionary, both in favorable and diverse acoustic conditions. An additional aspect of the project could involve creating adaptive forms that enable doctors to fill them out using voice input.

RESULTS:

1. Systematic literature review.
2. Familiarization with speech-to-text tools.
3. Review of available annotation tools.
4. Generation of medical vocabulary samples.
5. Testing available speech-to-text tools on recorded samples.
6. Familiarization with metrics for speech-to-text transcription.
7. Analysis of the obtained results of medical speech transcription using selected tools.
8. Creation and submission of a conference paper.

MAIN FEATURES, FUTURE WORKS:

Main features:

1. Selected speech-to-text creation tools:
- Whisper (small, medium, large)

TEAM PROJECT INFORMATION FOLDER – JUNE 2023

- Google
 - Microsoft Azure
2. Generation of 1200 samples of medical vocabulary (100 words) in mp3 format:
 - 600 female voices
 - 600 male voices
 - Including 100 syntetic male voices in 600 male voices and 100 female syntetic in 600 female voices
 3. Metrics used in result analysis:
 - Word Insertion Loss (WIL)
 - Word Error Rate (WER)
 - Character Error Rate (CER)
 - Mean Error Rate (MER)
 - Levenshtein distance
 - Jaccard distance Further
- Future work:**
1. Generating a larger amount of medical vocabulary samples.
 2. Retraining of the selected model.
 3. Testing trained model.
 4. Final Report preparation.