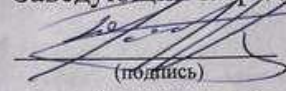


БЕЛОРУССКИЙ НАЦИОНАЛЬНЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
Факультет информационных технологий и робототехники
Кафедра «Программное обеспечение информационных систем и технологий»

ДОПУЩЕН К ЗАЩИТЕ

Заведующий кафедрой


(подпись)

Ю.В. Полозков
(инициалы и фамилия)

« 01 » 06 2023 г.

**РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА
ДИПЛОМНОГО ПРОЕКТА**

**«Программное средство для преобразования звукового сигнала в текст с
последующим его переводом»**

Специальность 1-40 05 01 «Информационные системы и технологии (по
направлениям)»
Направление специальности 1-40 05 01-04 «Информационные системы и технологии
(в обработке и представлении информации)»

Обучающийся
группы 10702219
(номер)

 18.05.23
(подпись, дата)

Н.П. Темерев

Руководитель

 29.05.23
(подпись, дата)

Ю.А. Скудняков

Консультанты:

по разделу «Компьютерное
проектирование»

 29.05.23
(подпись, дата)

Ю.А. Скудняков

по разделу «Охрана труда»

 19.05.23
(подпись, дата)

А.М. Лазаренков

по разделу «Экономика»

 25.05.23
(подпись, дата)

О.В. Куневич

Ответственный за нормоконтроль

 09/06.06.23
(подпись, дата)

Н.С. Домаренко

Объем проекта:
расчетно-пояснительная записка – 70 страниц;
графическая часть – 10 листов;
магнитные (цифровые) носители – 1 единиц.

РЕФЕРАТ

РАСПОЗНАВАНИЕ РЕЧИ, PYTHON, МАШИННОЕ ОБУЧЕНИЕ, НЕЙРОННЫЕ СЕТИ, СПЕКТРАЛЬНЫЙ АНАЛИЗ, РЕЧЬ-В-ТЕКСТ

Объектом исследования являются речевые сигналы с их спектральными характеристиками и текстовыми проекциями, а также алгоритмы осуществляющие преобразования речь-в-текст.

Цель проекта – разработка алгоритмов и программных средств для прямого преобразования речевых сигналов в текст с осуществлением перевода.

В процессе работы (проектирования) выполнены следующие исследования (разработки): проведено исследование алгоритмов анализа речевых сигналов, а также анализ существующих сервисов для перевода текстовых данных, разработан конвейер обработки данных, реализующий прямое преобразование речевых сигналов в текст с возможностью перевода, осуществлена интеграция программного пакета с интерфейсной частью.

Студент-дипломник подтверждает, что приведенный в дипломном проекте расчетно-аналитический материал объективно отражает состояние исследуемого процесса (разрабатываемого объекта), все заимствованные из литературных и других источников теоретические и методологические положения и концепции сопровождаются ссылками на их авторов.

Дипломный проект: 70 с., 10 рис., 15 табл., 10 чертежей, 27 источников, 1 прил.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 Николенко, С., Кадури, Е., Архангельская, Е. Глубокое обучение. -Издательство «Питер», 2019. – 259 с.
- 2 Официальная документация к библиотеке TensorFlow [Электронный ресурс]. – Режим доступа: https://www.tensorflow.org/api_docs
- 3 Официальная документация к фреймворку Tornado [Электронный ресурс]. – Режим доступа: <https://www.tornadoweb.org/en/stable/>
- 4 Coates, B. Huval, T. Wang, D. J. Wu, A. Y. Ng, and B. Catanzaro. Deep learning with COTS HPC. International Conference on Machine Learning, 2013.
- 5 Graves, S. Fernandez, F. Gomez, and J. Schmidhuber. Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks, InICML. – ACM, 2006. – pp. 369–376.
- 6 Aaron van den Oord, Sander Dieleman, Heiga Zen†, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, Koray Kavukcuoglu. WaveNet: a generative model for raw audio. – arxiv.org, September 19, 2016.
- 7 Alan W. Black. Perfect synthesis for all of the people all of the time. – IEEE TTS Workshop, 2002.
- 8 Aurélien Géron. Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. – 2019.
- 9 Awni Hannun, Carl Case, Jared Casper, Bryan Catanzaro, Greg Diamos, Erich Elsen, Ryan Prenger, Sanjeev Satheesh, Shubho Sengupta, Adam Coates, Andrew Y. Ng. Deep Speech: Scaling up end-to-endspeech recognition. – arxiv.org, December 19, 2014
- 10 Chenguang Lu. The CM Algorithm for the Maximum Mutual Information Classifications of Unseen Instances. – arxiv.org
- 11 D. Schwarz. Current research in Concatenative Sound Synthesis. – Proceedings of the International Computer Music Conference (ICMC), 2005.
- 12 D. W. Griffin and J. S. Lim. Signal estimation from modified short-time Fourier transform. – IEEE Transactions on Acoustics, Speech and Signal Processing, 1984. – pp. 236–243.
- 13 Eren Golge. Text to Speech Deep Learning Architectures. – February 20, 2018
[Электронный ресурс] <http://www.erogol.com/text-speech-deep-learning-architectures/>
- 14 George Roter. Sharing our Common Voices – Mozilla releases the largest to-date public domain transcribed voice dataset. – February 28, 2019
- 15 Jonathan Shen¹, Ruoming Pang¹, Ron J. Weiss¹, Mike Schuster¹, Navdeep Jaitly¹, Zongheng Yang, Zhifeng Chen¹, Yu Zhang¹, Yuxuan Wang¹, RJ Skerry-Ryan¹, Rif A. Saurous¹, Yannis Agiomyrgiannakis, and Yonghui Wu¹. Natural TTS synthesis by

- conditioning WaveNet on mel spectrogram predictions. – arxiv.org, February 16, 2018
- 16 Julia Zhang. Language Generation and Speech Synthesis in Dialogues for Language Learning. – Master’s thesis. – Section 5.6 on page 54.
- 17 Juri Ganitkevitch. Speaker Adaptation using Maximum Likelihood Linear Regression. – Seminar Automatic Speech Recognition. – summer term 2005
- 18 L.F. Lamel, J.L. Gauvain, B. Prouts, C. Bouhier, R. Boesch. Generation and Synthesis of Broadcast Messages. – Proceedings ESCA-NATO Workshop and Applications of Speech Technology, September 1993.
- 19 Lawrence R. Rabiner. A tutorial on Hidden Markov Models and selected applications in speech recognition. Proceedings of the IEEE, 77 (2) . – pp. 257–286.
- 20 M. Schuster and K. K. Paliwal. Bidirectional recurrent neural networks. – IEEE Transactions on Signal Processing, November 1997. Vol. 45, no. 11, – pp. 2673–2681.
- 21 R. Muralishankar, A.G. Ramakrishnan, P. Prathibha. Modification of Pitch using DCT in the Source Domain. – Speech Communication, vol. 42 (2), February 2004. – pp. 143–154.
- 22 Romain Serizel and Diego Giuliani. Vocal tract length normalization approaches to DNN-based children’s and adults’ speech recognition. – arxiv.org
- 23 S. Hochreiter and J. Schmidhuber. Long short-term memory. – Neural Computation, November 1997. vol. 9, no. 8, – pp. 1735–1780.
- 24 Sneha Lukose, Savitha S. Upadhyaya. Text to speech synthesizer-formant synthesis. - International Conference on Nascent Technologies in Engineering (ICNTE), 27 Jan 2017.
- 25 Stan Salvador, Philip Chan. FastDTW: Toward Accurate Dynamic Time Warping in Linear Time and Space. – KDD Workshop on Mining Temporal and Sequential Data, 2004. – pp. 70–80
- 26 Ye Jia, Ron Weiss. Introducing Translatotron: An End-to-End Speech-to-Speech Translation Model. – Google Inc. blog page – May 15, 2019
- 27 Yuxuan Wang, RJ Skerry-Ryan, Daisy Stanton, Yonghui Wu, Ron J. Weiss†, Navdeep Jaitly, Zongheng Yang, Ying Xiao, Zhifeng Chen, Samy Bengio†, Quoc Le, Yannis Agiomyrgiannakis, Rob Clark, Rif A. Saurous. Tacotron: towards end-to-end speech synthesis. arxiv.org, April 6, 2016