ANASTASIA KUZNETSOVA

Graduate Research Assistant | PhD student

@ anakuzne@iu.edu (812) 558-80-55 GitHub: ana-kuznetsova in linkedin.com/in/anastasia-kuznetsova-2bb66b116/

EDUCATION

PhD student in Computer Science and Computational Linguistics (3.54 GPA)

Indiana University

Aug 2019 – May 2024 Bloomington, IN, USA

MA in Computational Linguistics

NRU Higher School of Economics

Sep 2017 – Jun 2019 Moscow, Russia

EXPERIENCE

Machine Learning Engineer (Intern)

Rev.com

Jun 2021 – Aug 2021 Remote, Austin, TX, USA

Graduate Research Assistant

Indiana Univeristy Bloomington

Aug 2019 – Present Bloomington, IN, USA

Software Developer

MTS, Artificial Intelligence Department

Jan–Jul 2019 Moscow, Russia

Mentor

Google Summer of Code, Google Code-In, Apertium

2018, 2019, 2020 Remote

Student Participant

Google Summer of Code, Apertium

May - Aug, 2018 Remote

SKILLS

- Expertise: Speech Recognition, Speech enhancement, Deep Learning, Algorithm Design
- Languages: Python, R, C++
- Libraries: PyTorch, Torchaudio, Tensorflow, Fairseq, OpenAI gym.

PUBLICATIONS

- *Kuznetsova, A., A.Kumar, J. Drexler Fox, and Francis Tyers.* Curriculum Optimization for Low-resource Speech Recognition (ICASSP 2022, Submitted)
- *Vyas, P., Kuznetsova, A., Williamson, D.S.* (2021) Optimally Encoding Inductive Biases into the Transformer Improves End-to-End Speech Translation. Proc. Interspeech 2021, 2287-2291, doi: 10.21437/Interspeech.2021-2007
- *Kuznetsova,A. and F. Tyers.* A finite-state morphological analyser for Paraguayan Guaraní. Proceedings of the First Workshop on Natural Language Processing for Indigenous Languages of the Americas, p. 81 - 89.
- *Zueva, A., A. Kuznetsova, and F. Tyers.* "A finite-state morphological analyser for evenki." Proceedings of The 12th Language Resources and Evaluation Conference. 2020.

AWARDS

Best Student Paper Award at INTERSPEECH 2021

CURRENT PROJECTS

Low-resource ASR with language independent self-supervised representations

Researching the effect of resource rich language transfer learning on low-resource ASR systems. Building language-independent speech representations with contrastive predictive coding (CPC).

Contrastive Predictive Coding for Speech Enhancement

The project involves the usage of self-supervised audio representations along with human assessment scores and aims to improve signal quality and intellefibility.

A reinforcement-learning approach to curriculum generation for ASR

Applied curriculum learning approach and k-armed bandit algorithms to optimize end-to-end speech recognition system. The method mitigates the lack of training data and achieves 30% WER relative reduction.

PAST PROJECTS

Google Summer of Code: Machine Translation for Guarani–Spanish language pair

Built FST solution for rule-based low-resource machine translation system.

Speaker Identification system

Implemented RNN-based speaker ID system exploiting Siamese networks and speaker embeddings with one-shot learning approach.

BPE weighting of morphological analyser for Paraguayan Guarani

Weighting of the morphological analyser based on finite-state technology with Byte Pair encoding algorithm. **Morphological**

Disambiguation for Paraguayan Guaraní

Developed rule-based grammar (Constraint Grammar formalism).