

IRINA SAPARINA

Edinburgh, UK

I.Saparina@sms.ed.ac.uk ♦ saparina.github.io

PROFILE

I am a 3rd year Ph.D. student at the University of Edinburgh, under the supervision of Prof. Mirella Lapata. I work on Natural Language Processing and Understanding, with a research focus on Semantic Parsing and Code Generation, particularly in text-to-SQL semantic parsing. My primary area of interest is addressing linguistic variability and ambiguity.

EDUCATION

Ph.D. student, Informatics October 2022 - present
The University of Edinburgh *Edinburgh, UK*

Advisor: [Mirella Lapata](#)

Thesis: Linguistic Variability and Ambiguity in Semantic Parsing

Ph.D. student, Computer Science (GPA: 9.5/10.0) November 2020 - August 2022
HSE University *Moscow, Russia*

Advisor: [Anton Osokin](#)

Thesis: Augmenting Semantic Parsing with Intermediate Representations and Search

Master of Computer Science, with honours (GPA: 9.5/10.0) September 2018 - June 2020
Skolkovo Institute of Science and Technology and HSE University *Moscow, Russia*

Thesis: Cost-sensitive training for autoregressive models

Bachelor of Computer Science, with honours (GPA: 4.9/5.0) September 2014 - June 2018
Moscow State University *Moscow, Russia*

Thesis: Language-Independent Methods of Extracting Keywords in Web Texts

PUBLICATIONS

AMBROSIA: A Benchmark for Parsing Ambiguous Questions into Database Queries. [PDF](#), [Website](#)

Irina Saparina and Mirella Lapata.

We collected a new dataset for text-to-SQL semantic parsing with ambiguous questions and benchmarked LLMs to reveal their challenges in interpreting ambiguities.

Under review.

Improving Generalization in Semantic Parsing by Increasing Natural Language Variation. [PDF](#), [Code](#)

Irina Saparina and Mirella Lapata.

We proposed data augmentation to enhance the robustness of text-to-SQL parsers against natural language variations.

EACL 2024.

Searching for Better Database Queries. [PDF](#), [Code](#)

Anton Osokin and **Irina Saparina** and Ramil Yarullin.

We augmented semantic parsers with a search algorithm that looks for a query satisfying the task-specific criterion.

Findings of EACL 2023.

SPARQLing Database Queries from Intermediate Question Decompositions. [PDF](#), [Code](#)

Irina Saparina and Anton Osokin.

We unlocked training with simpler annotation using intermediate question representation (from the Break dataset) and achieved comparable execution accuracy with the state-of-the-art methods trained with full annotation.

EMNLP 2021.

Cost-Sensitive Training for Autoregressive Models. [PDF](#)

Irina Saparina and Anton Osokin.

We proposed new training losses for autoregressive models to avoid the disturbing properties of standard training and showed the quality improvement of machine translation, code generation and word ordering.

WiML workshop of NeurIPS 2019.

WORK EXPERIENCE

Yandex Research and HSE University Joint Laboratory <i>Research Assistant</i>	July 2020 - September 2022 <i>Moscow, Russia</i>
Proposed training with simpler annotation and inference with external criteria in the text-to-SQL semantic parsing.	
Samsung Laboratory, Centre of Deep Learning and Bayesian Methods <i>Research Assistant</i>	June 2018 – July 2020 <i>Moscow, Russia</i>
Investigated imitation learning algorithms for training sequence prediction models.	
Laboratory of Programming Technologies, Moscow State University <i>Intern</i>	June 2017 – September 2017 <i>Moscow, Russia</i>
Developed a Python tool for analyzing texts posted on social media.	

AWARDS

Scholarship of Advanced Doctoral Programme at HSE University	2020 - 2022
EMNLP Student Travel Grant	2021
NeurIPS Travel Grant	2019
Russian Academic Excellence Scholarship	2014-2020
Winner of SAS Data Analysis Hackathon	2017

SERVICE

Reviewer: WiML workshop of NeurIPS 2019, EMNLP 2023, ACL ARR 2024 February (ACL 2024), June (EMNLP 2024)

ADDITIONAL PROJECTS

Compressed Transformer. Developed a method for compressing Transformer using tensor decomposition ([code](#)).
AdaShift. Reproduced the failure cases of Adam and re-implemented the AdaShift optimizer that fixed them ([code](#)).
Neural ODE. Re-implemented the Neural ODE training and applied it to the text classification task ([code](#)).

SKILLS

- Python (daily use), SQL, C/C++
- PyTorch, HF Transformers, Git, Docker, Linux, Slurm, Kubernetes, LaTeX
- English (Upper-Intermediate), Russian (Native)

TEACHING

Deep Learning , HSE University	
Lecturer	Spring 2022
Seminar Teacher	Fall 2020
Teaching Assistant	Spring and Fall 2019
Bayesian Methods of Machine Learning , HSE University and Moscow State University	
Teaching Assistant	Fall 2018
Summer School on Deep Learning and Bayesian Methods (Deep Bayes)	
Assistant	Summer 2018, Summer 2019

WINTER AND SUMMER SCHOOLS

Statistics, Artificial Intelligence, Machine Learning, Probability, Learning Theory Event	Fall 2021
Math Of Machine Learning	Winter 2019-2020
Lisbon Machine Learning School (LxMLS)	Summer 2019