IRINA SAPARINA

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 in Informatics, The University of Edinburgh　　　　　　　　　　October 2022 - present
Thesis: Linguistic Variability and Ambiguity in Semantic Parsing　　　　　　　　　　*Edinburgh, UK*

Ph.D. student in Computer Science, HSE University (GPA: 9.5/10.0)　　November 2020 - August 2022
Thesis: Augmenting Semantic Parsing with Intermediate Representations and Search　　*Moscow, Russia*

Master of Computer Science, HSE University (GPA: 9.5/10.0)　　September 2018 - June 2020
Thesis: Cost-sensitive training for autoregressive models　　　　　　　　　　*Moscow, Russia*

Bachelor of Computer Science, Moscow State University (GPA: 4.9/5.0)　　September 2014 - June 2018
Thesis: Language-Independent Methods of Extracting Keywords in Web Texts　　　　*Moscow, Russia*

WORK EXPERIENCE

Yandex Research, Research Assistant　　　　　　　　　　July 2020 - September 2022
Improved training and inference in text-to-SQL semantic parsing using auxiliary annotation.　　*Moscow, Russia*

Samsung Laboratory, Research Assistant　　　　　　　　　　June 2018 – July 2020
Developed imitation learning algorithms for training sequence prediction models.　　*Moscow, Russia*

Laboratory of Programming Technologies, Moscow State University, Intern　　June 2017 – September 2017
Developed a Python tool for analyzing texts posted on social media.　　　　*Moscow, Russia*

PUBLICATIONS

𝔸𝕄𝔹ℝ𝕆𝕊𝕀𝔸**: A Benchmark for Parsing Ambiguous Questions into Database Queries.**　　PDF, Website
Irina Saparina and Mirella Lapata.
NeurIPS 2024 (Spotlight).

Improving Generalization in Semantic Parsing by Increasing Natural Language Variation.　　PDF, Code
Irina Saparina and Mirella Lapata.
EACL 2024 (Oral).

Searching for Better Database Queries.　　PDF, Code
Anton Osokin and **Irina Saparina** and Ramil Yarullin.
Findings of EACL 2023.

SPARQLing Database Queries from Intermediate Question Decompositions.　　PDF, Code
Irina Saparina and Anton Osokin.
EMNLP 2021.

Cost-Sensitive Training for Autoregressive Models.　　PDF
Irina Saparina and Anton Osokin.
WiML @ NeurIPS 2019.

SKILLS

- Python (daily use), SQL, C/C++
- PyTorch, HF Transformers, Git, Docker, Linux, Slurm, Kubernetes, LaTeX
- Reviewer: WiML @ NeurIPS 2019, EMNLP 2023, ACL ARR 2024 February (ACL 2024), June (EMNLP 2024)