Anton BELYY

anton.belyy@gmail.com · avbelyy.github.io

Scholar · GitHub · LinkedIn

EDUCATION

SEP '19 — PRESENT	**PhD in Computer Science (NLP)**, Johns Hopkins University, USA Advisors: Benjamin Van Durme (Primary), Vladimir Braverman (Secondary) • **M.Sc.Eng.** degree expected in May 2022. GPA: 4.0
SEP '13 — MAY '18	**B.Sc. in Informatics and Applied Mathematics**, ITMO University, Russia Thesis: Construction and Quality Evaluation of Heterogeneous Hierarchical Topic Models • **Distinguished thesis award** (given to 2 out of approx. 50 undergraduates)

EMPLOYMENT

JUN '21 — AUG '21	**Research Intern**, Microsoft Semantic Machines, USA • Built guided annotation interface to help label semantic parsing data 35% faster [preprint]

SEP '19 – PRESENT	**Graduate Research Assistant**, Johns Hopkins University, USA
	 • **Semantic data mining**: proposed novel ARM-based algorithm for script induction [paper],
	   built SchemaBlocks, Scratch-like annotation interface for complex event scenarios [preprint]
	 • **Knowledge base completion**: building a human-in-the-loop system to "complete" Wikidata
	   based on triples inferred from Wikipedia list articles and verified by an expert *(current project)*
	 • **Text generation**: built demo for InFillmore, our FrameNet frame-guided NLG model [paper]

SEP '17 — AUG '19	**Senior Data Scientist** in **Compliance Risks** and **AI lab**, Tochka Bank, Russia • **Risk scoring**: vectorized new data sources for 200K+ bank clients and 50M+ transactions, generated temporal/spatial features using Hadoop/Spark to improve scoring accuracy by 10% • **Communication analysis**: built intent recognition models to classify 90% customer inquiries • **Call center planning**: using OR-Tools, automated CC planning while improve accuracy by 10% • **ML culture**: interviewed and mentored 3 junior ML engineers, designed internal ML guidelines
OCT '16 — MAY '18	**Undergraduate Research Assistant** (part-time), ITMO University, Russia • **Exploratory search via topic models**: built framework for document-level semantic search using interpretable TMs [1, 2, 3]. Built system demo using MongoDB, ZeroMQ, and FoamTree • **Plagiarism detection**: proposed novel evaluation framework [paper] and ML model [paper]
OCT '15 — OCT '16	**Software Engineer** in **Ads**, VK.com, Russia • **URL fraud**: built service to periodically detect malicious URL redirect changes in VK ads • **Click fraud**: built ML models to detect users that generate fraudulent clicks in VK ad network. Model was deployed semi-automatically and helped recover up to 3% monthly ad revenue • **Ads search**: launched moderator search interface (incl. full-text search) over 30M+ VK ads • **Ads scoring**: implemented advertiser ranking for faster moderation of top-10% clients
JUL '15 — SEP '15	**Software Engineer Intern** in **.NET tools**, JetBrains, Russia • Implemented methods for automatic memory leak detection in the dotMemory profiler

TEACHING EXPERIENCE

JAN '21 — MAY '21	**Introduction to Algorithms 601.433/633 (Head TA)**, JHU (100+ students) • Managed 9 CAs and 1 TA, created homework and exam problems, held weekly office hours
SEP '17 — AUG '19	**Natural Language Processing (TA)**, Coursera (40,000+ students by Sep '19) • Answered 200+ students' questions, helped create homework and project assignments

LANGUAGES AND TECHNOLOGIES

LANGUAGES	Python (proficient); JavaScript, bash (intermediate); C#, C++, Haskell, x86 assembly (coursework)
TECHNOLOGIES	pandas, sklearn, XGBoost, pytorch, faiss, nmslib; Docker, *SQL, MongoDB, Lucene, Hadoop/Spark