

Anton BELY

Scholar | [GitHub](#) | [LinkedIn](#)

abel@jhu.edu

EDUCATION

AUG '19 — MAY '25	PhD in Computer Science (NLP) , Johns Hopkins University Interests: Large-scale NLP, Explainable AI / ML, Indexing & Sketching for NLP Advisors: Benjamin Van Durme (Primary), Vladimir Braverman (Co-advisor)
SEP '13 — JUN '18	B.Sc. in Informatics and Applied Mathematics , ITMO University Thesis: Construction and Quality Evaluation of Heterogeneous Hierarchical Topic Models [draft] Advisors: Andrey Filchenkov (Primary), Konstantin Vorontsov (Co-advisor) Grade: 5 / 5 (with a special distinction)

RESEARCH EXPERIENCE

AUG '19 — ONGOING	Research Assistant , Johns Hopkins University, USA Supervisors: Benjamin Van Durme, Vladimir Braverman <ul style="list-style-type: none">• Association rule based schema induction and prediction• Approximate near neighbor search for large-scale multilabel classification
MAR '18, JUN '18	Visiting Student , Université Grenoble Alpes, France Supervisors: Massih-Reza Amini, Yury Maximov <ul style="list-style-type: none">• Extreme multi-class classification using Pegasos algorithm [draft]• Sparsification for kernelized SVM using Metropolis-Hastings sampling
SEP '17 — FEB '18	Research Assistant , ITMO University, Russia Supervisor: Andrey Filchenkov <ul style="list-style-type: none">• Novel evaluation metric for external plagiarism detection [paper]• Framework for external plagiarism detection applied for Russian language corpora [paper]
MAR '17 — AUG '17	Research Assistant , Russian Academy of Sciences, Russia Supervisor: Konstantin Vorontsov <ul style="list-style-type: none">• Hierarchical topic modeling for exploratory search over heterogeneous sources [paper]• Topic-model driven exploratory search engine (backend, frontend, devops scripts) [code]
JUL '14 — MAR '15	R&D Intern , Synopsys, Russia Supervisor: Sergey Yakushkin <ul style="list-style-type: none">• Static memory allocation for micro-controllers using graph k-coloring methods• Implementation and evaluation of three coloring algorithms, technical report with results

WORK EXPERIENCE

SEP '17 — AUG '19	Senior Data Scientist , Tochka Bank, Russia <ul style="list-style-type: none">• Predictive modeling for anti-money laundering compliance service• Exploratory analysis of client communication (chat and phone call sessions)
SEP '15 — SEP '16	Software Developer , VK.com, Russia <ul style="list-style-type: none">• Fraud click detection in the external VK advertisement network• Automatic pre-moderation of ads from trustworthy advertisers
JUL '15 — SEP '15	Software Developer Intern , JetBrains, Russia <ul style="list-style-type: none">• Implementation of a garbage collection algorithm for dotMemory profiler

TEACHING AND ACADEMIC SERVICE

Nov '19	Volunteer (FOCS 2019) http://focs2019.cs.jhu.edu
JUN '18 — JUN '20	PC member (AINL 2018 — 2020) https://ainlconf.ru
SEP '17 — AUG '19	Teaching Assistant (Natural Language Processing) https://coursera.org/learn/language-processing

PUBLICATIONS

1. Improved Evaluation Framework for Complex Plagiarism Detection.

Belyy, A., Dubova, M., & Nekrasov, D. (2018). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL)*, Vol. 2, pp. 157-162. [\[paper\]](#) [\[poster\]](#) [\[code\]](#)

2. Quality Evaluation and Improvement for Hierarchical Topic Modeling.

Belyy, A., Seleznova, M., Sholokhov, A., & Vorontsov, K. (2018). In *Proceedings of the 24th International Conference on Computational Linguistics and Intellectual Technologies (Dialogue)*, Issue 17, pp. 110-123. [\[paper\]](#) [\[slides\]](#)

3. Framework for Russian Plagiarism Detection Using Sentence Embedding Similarity and Negative Sampling.

Belyy, A., & Dubova, M. (2018). In *Proceedings of the 24th International Conference on Computational Linguistics and Intellectual Technologies (Dialogue)*, Issue 17, pp. 96-109. [\[paper\]](#) [\[slides\]](#) [\[code\]](#)

4. MEMOIR: Multi-class Extreme Classification with Inexact Margin (work in progress).

Belyy, A., Sholokhov, A., Amini, M.-R., & Maximov, Yu. [\[draft\]](#) [\[code\]](#)

CONFERENCES AND SCHOOLS

JUL '19	9th Lisbon Machine Learning School , Lisbon, Portugal Summer school on Machine Learning and Natural Language Processing
NOV '17	60th MIPT Scientific Conference , Moscow, Russia Presentation on aggregating text collections of heterogeneous web sources [slides] (best presentation award)
AUG '16	ISMW FRUCT Conference and School , St. Petersburg, Russia Summer school on Social Media Analysis and Topic Modeling

SKILLS AND TOOLS

SPOKEN LANGUAGES	Russian (Native Speaker), English (Full Professional Proficiency), German (Elementary)
PROGRAMMING	Python, Haskell, C / C++, Java, Assembler (with SIMD extensions), *nix shell, *SQL, PHP, HTML
TOOLS & LIBRARIES	Docker, git, LaTeX, Node.JS, Socket.IO, MongoDB, NumPy, SciPy, Pandas, BigARTM

COURSEWORK

GRADUATE	Natural Language Processing, Randomized Algorithms
UNDERGRADUATE	Algorithms and Data Structures, Discrete Mathematics, Computer Architecture, Computational Geometry, Calculus, Linear Algebra, Probability Theory, Statistics, Information Theory, Statistical Machine Learning, Mathematical Logic, Type Theory, Complexity Theory, Language & Automata Theory, Optimization Theory, Game Theory, Operating Systems, Parallel & Distributed Programming, C++, Java, Functional Programming, Assembler, Computer Networks, Databases