

Anton BELYI

abel@jhu.edu · avbelyy.github.io

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

EDUCATION

AUG '19 — PRESENT	PhD in Computer Science (NLP) , Johns Hopkins University, U.S.A. Advisors: Benjamin Van Durme (Primary), Vladimir Braverman (Secondary)
SEP '13 — JUN '18	B.Sc. in Informatics and Applied Mathematics , ITMO University, Russia Thesis: Construction and Quality Evaluation of Heterogeneous Hierarchical Topic Models Advisors: Andrey Filchenkov (Primary), Konstantin Vorontsov (Secondary)

RESEARCH INTERESTS

Large-scale information retrieval, structured text generation, streaming and approximation algorithms, ML Theory in NLP

RESEARCH EXPERIENCE

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 with Pegasos and MIPS algorithms [report]
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 in Russian [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 (full-stack implementation) [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 of three graph coloring algorithms, technical report with results

WORK EXPERIENCE

SEP '17 — AUG '19	Senior Data Scientist , Tochka Bank, Russia <ul style="list-style-type: none">• Anti-money laundering: prediction and scoring models• Client communications: intent recognition and exploratory analysis
SEP '15 — SEP '16	Software Developer , VK.com, Russia <ul style="list-style-type: none">• Fraud (anomaly) detection in ads' clicks• Classification of 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

PUBLICATIONS

1. Script Induction as Association Rule Mining.

Belyy, A., & Van Durme, B. (2020). In *Proceedings of the 1st Joint Workshop on Narrative Understanding, Storylines, and Events (NUSE)*, pp. 55-62. [\[paper\]](#) [\[slides\]](#) [\[code\]](#)

2. 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. [\[paper\]](#) [\[poster\]](#) [\[code\]](#)

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. 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. [\[paper\]](#) [\[slides\]](#)

REVIEWING

ICDT	2021 (secondary)
ICLR	2021 (secondary)
AKBC	2020 (secondary)
ACL	2020 (secondary)
AINL	2018, 2019, 2020

TEACHING EXPERIENCE

SEP '17 — AUG '19

Teaching Assistant (Natural Language Processing)

<https://coursera.org/learn/language-processing>

The most massive NLP course on Coursera as of Aug'19 ($\approx 30k$ active learners and ≈ 800 completers).

I beta-tested the course, wrote and graded assignments, answered students' questions on forum.

COURSEWORK

AI / ML	Natural Language Processing, Machine Learning, Optimization Theory
ALGORITHMS	Discrete Mathematics, Algorithms & Data Structures, Randomized Algorithms, Computational Geometry, Complexity Theory, Sketching and Indexing, Information Retrieval
MATHEMATICS	Linear Algebra, Calculus, Probability Theory, Intro Stats, Statistical Theory, Information Theory, Mathematical Logic, Type Theory, Language & Automata Theory, Game Theory
PROGRAMMING	C++, Java, Haskell, Assembler, Web Development, Mobile Development (Android)
SYSTEMS	Computer Architecture, Parallel & Distributed, Operating Systems, Networks, Databases

SKILLS AND TOOLS

SPOKEN LANGUAGES	Russian (Native Proficiency), English (Full Professional Proficiency), Chinese (Elementary)
PROGRAMMING	Python, Haskell, C / C++, Java, JavaScript, Assembler, *nix shell, *SQL, PHP, HTML
FRAMEWORKS	Research: NumPy, Pandas, PyTorch, fairseq, nmslib, BigARTM WebDev: Node.JS, Socket.IO, MongoDB, flask, memcache
TOOLS	Docker, git, LaTeX, (C)Make, JetBrains IDEs