# Anton BELYY

abel@jhu.edu · avbelyy.github.io

Scholar · GitHub · LinkedIn

## EDUCATION

| | |
|---|---|
| AUG '19 — PRESENT | **PhD in Computer Science (NLP)**, Johns Hopkins University, U.S.A.<br>Advisors: Benjamin Van Durme (Primary), Vladimir Braverman (Secondary) |
| SEP '13 — JUN '18 | **B.Sc. in Informatics and Applied Mathematics**, ITMO University, Russia<br>Thesis: Construction and Quality Evaluation of Heterogeneous Hierarchical Topic Models<br>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<br>Supervisors: Massih-Reza Amini, Yury Maximov<br>• Extreme multi-class classification with Pegasos and MIPS algorithms [report] |
| SEP '17 — FEB '18 | **Research Assistant**, ITMO University, Russia<br>Supervisor: Andrey Filchenkov<br>• Novel evaluation metric for external plagiarism detection [paper]<br>• Framework for external plagiarism detection in Russian [paper] |
| MAR '17 — AUG '17 | **Research Assistant**, Russian Academy of Sciences, Russia<br>Supervisor: Konstantin Vorontsov<br>• Hierarchical topic modeling for exploratory search over heterogeneous sources [paper]<br>• Topic-model driven exploratory search engine (full-stack implementation) [code] |
| JUL '14 — MAR '15 | **R&D Intern**, Synopsys, Russia<br>Supervisor: Sergey Yakushkin<br>• Static memory allocation for micro-controllers using graph $k$-coloring methods<br>• Implementation of three graph coloring algorithms, technical report with results |

## WORK EXPERIENCE

| | |
|---|---|
| SEP '17 — AUG '19 | **Senior Data Scientist**, Tochka Bank, Russia<br>• Anti-money laundering: prediction and scoring models<br>• Client communications: intent recognition and exploratory analysis |
| SEP '15 — SEP '16 | **Software Developer**, VK.com, Russia<br>• Fraud (anomaly) detection in ads' clicks<br>• Classification of trustworthy advertisers |
| JUL '15 — SEP '15 | **Software Developer Intern**, JetBrains, Russia<br>• 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]

5. **MEMOIR: Multi-class Extreme Classification with Inexact Margin**.

   Belyy, A., Sholokhov, A. (2018) [preprint]

## 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 $\approx$ 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 |