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. 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

- Similarity search and information retrieval
- Human assisted data collection and curation
- Schema induction and structured text generation

Publications and Preprints

1. **InFillmore: Frame-Guided Language Generation with Bidirectional Context.**

   Ou, J., Weir, N., Belyy, A., Yu, F., & Van Durme, B. (2021). In *Proceedings of the 10th Conference on Lexical and Computational Semantics*, pp. 129-142. [paper] [poster] [talk] [demo]

2. **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] [talk] [code]

3. **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]

4. **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]

5. **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]

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

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

Work Experience

Sep '17 — Aug '19 | **Senior Data Scientist**, Tochka Bank, Russia
- Anti-money laundering: prediction and scoring models
- Client communications: dialog intent recognition and exploratory analysis

Sep '15 — Sep '16 | **Software Developer**, VK.com, Russia
- Fraud (anomaly) detection in ads' clicks
- Classification of trustworthy advertisers

Jul '15 — Sep '15 | **Software Developer Intern**, JetBrains, Russia
- Implementation of a garbage collection algorithm for dotMemory profiler

Research Experience

Jun '21 – Aug '21 | **Research Intern**, Microsoft Semantic Machines, U.S.A.

Supervisor: Charles Chen
- Guided K-best selection method for rapid prototyping of complex NLG models

Mar '18 – Jun '18 | **Visiting Student**, Université Grenoble Alpes, France

Supervisors: Massih-Reza Amini, Yury Maximov
- Extreme multi-class classification with Pegasos and MIPS algorithms [report]

Sep '17 — Feb '18 | **Research Assistant**, ITMO University, Russia

Supervisor: Andrey Filchenkov
- 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
- Hierarchical topic modeling for exploratory search over heterogeneous sources [paper]
- Topic-model driven exploratory search engine (full-stack implementation) [code]

Jul '14 — Mar '15 | **Research Intern**, Synopsys, Russia

Supervisor: Sergey Yakushkin
- Static memory allocation for micro-controllers using graph k-coloring methods
- Implementation of three graph coloring algorithms, technical report with results

Teaching Experience

Spring 2021 | **Intro Algorithms 601.433/633 (Head TA)**, JHU

Sep '17 — Aug '19 | **Natural Language Processing (TA)**, Online
https://coursera.org/learn/language-processing

Skills and Tools

Spoken languages	Russian (Native Proficiency), English (Bilingual Proficiency), Chinese (Elementary)
Programming	Python, C / C++, Haskell, Java, JavaScript, Assembler, *nix shell, *SQL, PHP, HTML
Frameworks	Research: NumPy, Pandas, ScikitLearn, PyTorch, fairseq, FAISS, nmslib, BigARTM
	WebDev: TypeScript, Node.JS, Socket.IO, MongoDB, flask, memcache
Tools	Docker, git, LaTeX, (C)Make, JetBrains IDEs