Anton M. Alekseev

Scopus: 57209096106
Researcher ID: AAE-1183-2020
ORCID: 0000-0001-6456-3329
Scholar: bit.ly/gsch-alekseev

Geo: St. Petersburg, Russia
GMail: anton.m.alexeyev
homepage: alexeyev.github.io
github: alexeyev

Interests natural language processing, deep learning, ml4se, digital humanities

Education

Computer Science Center, St. Petersburg, Russia (grad. 2014)
Alumnus, two tracks: Data Science & Software Engineering

St. Petersburg University, St. Petersburg, Russia (grad. 2014, *cum laude*)
Specialist, the Faculty of Mathematics &Mechanics,
Applied Informatics in the field of International Relations,

MOOCs
Social Network Analysis, UMich@Coursera, 2013
Functional Programming Principles in Scala, EPFL@Coursera, 2013

Also: PhD studies at AU RAS, PDMI RAS, and KSTU n. a. I. Razzakov.

Teaching

Teaching assistant/Lecturer **MCS SPbU**
Feb 2021 – up to now St. Petersburg, Russia
Workshops for the courses of PGMs & DL (pytorch).
Lectures on NLProc (in Russian).

Lecturer **Computer Science Center**
Sep 2017 – Feb 2022 St. Petersburg, Russia
Lectures on Natural Language Processing in Russian.

- Topics: from string processing algorithms to *Transformer*-based models, two lectures a week.
- The design of tests and homeworks (with D. Bobrovnikov and G. Rozhkov).
- The course's syllabus, slides and videorecords are available online.
  Also: YouTube playlist, 22 videos

Visiting Lecturer **ITMO University**
Feb 2018 – May 2019 St. Petersburg, Russia
Lectures on Natural Language Processing in English.

- Topics: necessary mathematical and algorithmic toolkits, string processing, **machine learning in NLProc**: language modeling, text classification, duplicate detection, text clustering and topic modeling, modern distributional semantics, POS-tagging, named entities recognition, syntax parsing and statistical machine translation elements.
- The design of tests, homeworks (with K. Buraya) and written examination.

Visiting Teacher **Higher School of Economics**
Sep 2016 – Dec 2016 St. Petersburg, Russia
A short-term contract for teaching a programming workshop in Russian. Introduction into **Programming and Data Analysis** in Specialized Environments.

- Topics: introduction into Python [main part of the course], **NumPy, scipy.sparse matrices, basic Pandas**, regular expressions, parsing XML, CSV, JSON, introduction into **machine learning with scikit-learn**, GNU coreutils, a peek at .git/.hg
- Developed a new and an up-to-date introductory programming course programme: course site @ HSE.ru, course materials.

Other experience

Academic Consultant (part-time) — **St. Petersburg Research Center**
Mar 2024 – up to now — St. Petersburg, Russia
Machine learning in applied research.

Acting Junior Researcher — **St. Petersburg Department of Steklov Mathematical Institute (PDMI RAS)**
Mar 2018 – up to now — St. Petersburg, Russia
AI Lab., led by Dr. Sergey Nikolenko →
→ Lab. of Appl. Prob. and Alg. Methods, led by Dr. Alexander Kulikov
Text-based recommender systems, information extraction, representation learning, user profiling, algorithms for packet classifiers, machine learning for various domains.

Machine Learning Specialist — **Native Media LLC**
Oct 2016 – Feb 2018 — St. Petersburg, Russia
Predictive services for native video advertising **nativeroll.tv**, from scratch.

Intelligent Systems Engineer — **SofIT Labs LLC**
Aug 2015 – Sep 2016 — St. Petersburg, Russia
NLProc and information retrieval for customer-assisting services.

Intern → Software Engineer — **Yandex**
Feb 2013 –Aug 2015 — St. Petersburg, Russia
Online reviews processing backend, then Yandex.Rabota.

Skills

Languages: Russian (Native), English (Advanced)
Programming Languages: Python, Scala, Java, bash, R (basic)
Tools: scikit-learn, keras, PyTorch, scipy, numpy, pandas, gensim
Online Classroom Platforms:
Stepik (online tests), repl.it (workshops, exam tests), Kaggle/Kaggle inClass
Other: LaTeX, HTML, CSS, Microsoft Office (World, Excel, PowerPoint)

Schools

Bayesian Methods in Deep Learning Summer School
Aug 26-30, 2017, Moscow, Russia.

The 11th Russian Summer School in Information Retrieval
Aug 21-25, 2017, Yekaterinburg, Russia.

DeepHack.Turing Summer school-hackathon (**Kaggle**-based participants selection)
Jul 24-30, 2017, Dolgoprudnyy, Russia.

5th Lisbon Machine Learning School.
Jul 16-23, 2015, Lisbon. Portugal.

International Social Network Analysis Summer School (HSE)
Aug 8–13, 2014, St. Petersburg, Russia.

The 44-th International Youth School-Conference "Current problems in mathematics"
Jan 27 - Feb 2, 2013, Yekaterinburg

The 6th Russian Summer School in Information Retrieval.
Aug 6-10, 2012, Yaroslavl, Russia.

Talks
(published conference papers presentations are not included)

Open (?) directions of research in Kyrgyz NLP [Russian]
(summer 2023)

Open Information Extraction survey: key papers, tools and datasets [Russian]
(summer 2021, ML Network, 4 Aug. 2021)

Same Words, Different Tone:
Genre-Specific Sentiment Lexicons for Digital Music Reviews [English]
(winter 2019, LMAC-III, 17 Dec. 2019)

Unlemmatization: Recovering Word Forms in Morph. Rich Languages [English]
(autumn 2017, AINL: St. Petersburg, Russia, 20-23 Sept. 2017)

Natural Language Processing: Introduction [Russian]
(spring 2017, Kaggle Club)

How to build Yandex/Google from scratch? [Russian]
(summer 2016, Futurum STREAM Camp)

Service

Reviewing for conferences: AAAI, CHI, AIST, CompLing International Workshop, LKE Symposium, SEIM.

Reviewing SPbU, NRU HSE & ITMO University students' BSc/MSc theses.

Twice a member of the attestation commission at the MSc dissertations defences (SPbU, Faculty of Philology, Math. Linguistics Department).

Scientific (co)advisor/consultant of
...Anastasia Predelina, B.Sc. in Math. and Comp. Science at St. Petersburg University.
Neural networks for coordination analysis (2022-2023; published in Doklady Mathematics)

...Georgii Angeni, B.Sc. in Math. and Comp. Science at St. Petersburg University.
Finding heuristic rules for solving the Boolean satisfiability problem using machine learning methods (2022-2023)

...Mikhail Shirokikh, M.Sc. in Mathematics at St. Petersburg University.
Efficient neural architectures for NP-hard problems and recommender systems
(2023-2024; accepted to ACM SIGIR conf. and the Journal of Math. Sciences)

*

Research papers

K. Khrabrov, A. Ber, A. Tsypin, K. Ushenin, E. Rumiantsev, A. Telepov, D. Protasov, I. Shenbin, A. Alekseev, M. Shirokikh, S. Nikolenko, E. Tutubalina, and A. Kadurin. ∇^2 dft: A universal quantum chemistry dataset of drug-like molecules and a benchmark for neural network potentials. *Accepted to NeurIPS D&B Track 2024 as a poster*, 2024.

M. Shirokikh, I. Shenbin, A. Alekseev, A. Volodkevich, A. Vasilev, A. V. Savchenko, and S. Nikolenko. Neural click models for recommender systems. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2553–2558, 2024.

M. Shirokikh, I. Shenbin, A. Alekseev, A. Volodkevich, A. Vasilev, and S. I. Nikolenko. User response modeling in recommender systems: A survey. *Zapiski Nauchnykh Seminarov POMI*, 530:141–190, 2023a.

A. Alekseev, A. V. Savchenko, E. Tutubalina, E. Myasnikov, and S. I. Nikolenko. Blending of predictions boosts understanding for multimodal advertisements. *Zapiski Nauchnykh Seminarov POMI*, 529:176–196, 2023a.

А. И. Пределина, С. Ю. Дуликов, and А. М. Алексеев. Нейросетевые методы выделения сочинительных связей. *Доклады Российской академии наук. Математика, информатика, процессы управления*, 514 (2):289–296, 2023.

A. Alekseev and G. Kabaeva. Hj-ky-0.1: an evaluation dataset for kyrgyz word embeddings. *Herald of KSTU*, 68 (4):1806–1814, 2023.

A. Alekseev, S. I. Nikolenko, and G. Kabaeva. Benchmarking multilabel topic classification in the kyrgyz language. In *International Conference on Analysis of Images, Social Networks and Texts (to appear in print)*. Springer, Cham, 2023b.

M. Shirokikh, I. Shenbin, A. Alekseev, and S. I. Nikolenko. Machine learning for sat: Restricted heuristics and new graph representations. *arXiv preprint arXiv:2307.09141*, 2023b.

A. Nesterov, G. Zubkova, Z. Miftahutdinov, V. Kokh, E. Tutubalina, A. Shelmanov, A. Alekseev, M. Avetisian, A. Chertok, and S. I. Nikolenko. Ruccon: Clinical concept normalization in russian. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 239–245, 2022.

A. Alekseev, Z. Miftahutdinov, E. Tutubalina, A. Shelmanov, V. Ivanov, V. Kokh, A. Nesterov, M. Avetisian, A. Chertok, and S. I. Nikolenko. Medical crossing: a cross-lingual evaluation of clinical entity linking. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 4212–4220, 2022a.

M. Vasilkovsky, A. Alekseev, V. Malykh, I. Shenbin, E. Tutubalina, D. Salikhov, M. Stepnov, A. Chertok, and S. I. Nikolenko. Detie: Multilingual open information extraction inspired by object detection. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence*, 2022.

Kuzma Khrabrov, Ilya Shenbin, Alexander Ryabov, Artem Tsypin, Alexander Telepov, Anton Alekseev, Alexander Grishin, Pavel Strashnov, Petr Zhilyaev, Sergey Nikolenko, and Artur Kadurin. nabladft: Large-scale conformational energy and hamiltonian prediction benchmark and dataset. *Physical Chemistry Chemical Physics*, 24(42):25853–25863, 2022.

A. Alekseev, E. Tutubalina, S. Kwon, and S. I. Nikolenko. Near-zero-shot suggestion mining with a little help from wordnet. In *International Conference on Analysis of Images, Social Networks and Texts*, pages 23–36. Springer, Cham, 2022b.

A. M. Alekseev and S. I. Nikolenko. Recovering word forms by context for morphologically rich languages. *Zapiski nauchnykh seminarov POMI*, 499(0):129–136, 2021.

R. B. Galinskii, A. M. Alekseev, and S. I. Nikolenko. Word-based russian text augmentation for character-level models. *Zapiski nauchnykh seminarov POMI*, 499(0):206–221, 2021.

A. Savchenko, A. Alekseev, S. Kwon, E. Tutubalina, E. Myasnikov, and S. Nikolenko. Ad lingua: Text classification improves symbolism prediction in image advertisements. In *Proceedings of the 28th International Conference on Computational Linguistics (COLING 2020)*, pages 1886–1892, 2020.

A. Alekseev and S. Nikolenko. Recognizing preferred grammatical gender in russian anonymous online confessions. In Petr Sojka, Ivan Kopeček, Karel Pala, and Aleš Horák, editors, *Text, Speech, and Dialogue*, pages 222–230, Cham, 2020. Springer International Publishing. ISBN 978-3-030-58323-1.

A. Alekseev, E. Tutubalina, V. Malykh, and S. Nikolenko. Improving unsupervised neural aspect extraction for online discussions using out-of-domain classification. *Journal of Intelligent & Fuzzy Systems*, 39(2):2487–2496, 2020.

I. Shenbin, A. Alekseev, E. Tutubalina, V. Malykh, and S. I. Nikolenko. Recvae: A new variational autoencoder for top-n recommendations with implicit feedback. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, WSDM '20, page 528–536, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450368223. doi: 10.1145/3336191.3371831. URL https://doi.org/10.1145/3336191.3371831.

V. Malykh, A. Alekseev, E. Tutubalina, I. Shenbin, and S. Nikolenko. Wear the right head: Comparing strategies for encoding sentences for aspect extraction. In *Analysis of Images, Social Networks and Texts: 8th International Conference, AIST 2019, Kazan, Russia, July 17–19, 2019, Revised Selected Papers 8*, pages 166–178. Springer, 2019.

S. I. Nikolenko, E. Tutubalina, V. Malykh, I. Shenbin, and A. Alekseev. Aspera: Aspect-based rating prediction model. In *Advances in Information Retrieval: 41st European Conference on IR Research, ECIR 2019, Cologne, Germany, April 14–18, 2019, Proceedings, Part II 41*, pages 163–171. Springer, 2019.

A. Alekseev and S. Nikolenko. User profiling in text-based recommender systems based on distributed word representations. *Communications in Computer and Information Science*, 661:196–207, 2017a. doi: 10.1007/978-3-319-52920-2_19.

A. Menshikova, D. Maglevanaya, M. Kuleva, S. Bogdanova, and A. Alekseev. Art critics and art producers: Interaction through the text. In *Digital Transformation and Global Society: Third International Conference, DTGS 2018, St. Petersburg, Russia, May 30–June 2, 2018, Revised Selected Papers, Part II 3*, pages 113–124. Springer, 2018.

A. Alekseev and S. Nikolenko. Word embeddings for user profiling in online social networks. *Computacion y Sistemas*, 21(2):203–226, 2017b. doi: 10.13053/CyS-21-2-2734.

A. Alekseev and S. I. Nikolenko. Predicting the age of social network users from user-generated texts with word embeddings. *Proceedings of the AINL FRUCT 2016 Conference*, 2017c.

R. Galinsky, A. Alekseev, and S. I. Nikolenko. Improving neural network models for natural language processing in russian with synonyms. *Proceedings of the AINL FRUCT 2016 Conference*, 2017.

Алексеев А.М. Тулупьев А.Л., Фильченков А. А. Декодирование последовательности состояний бинарной скрытой марковской модели, представленной в виде алгебраической байесовской сети, по последовательности наблюдений. *Труды СПИИРАН*, 1(24):165–177, 2013.

А. М. Алексеев. Автоматизация анализа популярности технологических областей в корпусе текстов русскоязычных электронных медиа на основе данных Википедии. In *Всероссийская научная конференция по проблемам информатики СПИСОК-2014*, pages 559–564, 2014.

Алексеев А.М. Методы построения и обработки скрытых марковских моделей, представленных в виде алгебраических байесовских сетей. In *СПИСОК-2013: Материалы всероссийской научной конференции по проблемам информатики (23–26 апреля 2013 г., Санкт-Петербург). СПб.: ВВМ*, pages 744–749, 2013.

Алексеев А.М. Представление и визуализация моделей для вычислительных экспериментов с локальными структурами алгебраических байесовских сетей. In *XV Международная конференция по мягким вычислениям и измерениям SCM'12 (25–27 Июня, Санкт-Петербург). Сборник докладов.*, volume 1, pages 215–219, 2012.

Тулупьев А.Л. Алексеев А.М., Фильченков А.А. Оценка правдоподобия наблюдений и декодирование скрытых состояний для скрытых марковских моделей на основе алгебраических байесовских сетей. In *Региональная информатика-2012 (РИ-2012). XIII Санкт-Петербургская международная конференция. Санкт-Петербург, 24—26 октября, 2012 г.: Материалы конференции.*, page 28, 2012.