Ianina Anastasia

Contact Information

- **Address:** 28a, Pervomaiskaya St, Dolgoprudny, Moscow Region, Russia, 141700
- **E-mail:** yanina.anastasia.mipt@gmail.com
- **Telephone:** +7-925-309-59-03
- **Skype:** yanina_anastasia

Education

Samara Lyceum of Information Technologies and Computer Science
Samara, Russia (Sep 2005 – June 2012)

- Secondary and high school education, GPA: 5.00/5.00

Moscow Institute of Physics and Technology (State University)
Department of Innovations and High Technology (Sep 2012 to present)

- BSc in Applied Mathematics and Computer Science (Sep 2012 – July 2016)
  GPA: 9.3/10.0
  Bachelor thesis: Multimodal topic modeling or articles from collective blog for exploratory search. Scientific supervisor: Konstantin Vorontsov.
- MSc in Applied Mathematics and Computer Science (Sep 2016 to present)
  GPA: 10.0/10.0

The Yandex School of Data Analysis
Big Data department (Sep 2014 to May 2016)

- GPA: 4.9/5.0
- Area of study: machine learning, data science, artificial neural networks, Bayesian methods, probability theory, stochastic processes, image recognition, analytics, programming in C++, Python, R, distributed computing.

Work experience

Jul 2014 – Aug 2014
ABBYY
Intern in ABBYY Labs.

- Responsibilities: developing a GUI application “Math Hero” for recognition symbols in mathematical formulas (C#). Designing and implementing a character recognition system, developing a service for marking up mathematical symbols manually.

Feb 2016 – to present
Yandex
Analyst in geo-search team

- Responsibilities: performing analysis on big data, designing and implementing reporting dashboards with main metrics for geo-search services (Python), working with logs and detecting anomalies in data (Python, MapReduce, Hadoop, SQL), conducting AB-experiments, developing and implementing AB-metrics (C++), developing features for geo-search services (Python, C++), performing feature selection, developing and training ranking formulas, teaching interns and new members in the team.

Knowledge and experience

- Python – 4 years
- C/C++ – 3 years
- Java, C#, R - basic
- Machine learning and data science – advanced
- MapReduce & Hadoop – advanced
- SQL – good
- Mathematical statistics, probability theory, stochastic processes - good
- Decent knowledge of data structures: arrays, lists, hash tables, heaps (binary, binomial, fibonacci), trees (RB-Tree, B-Tree, Kd-Tree). Good knowledge of algorithms: sorting, traversal, string pattern search, graph algorithms
- Multithreading programming – understanding the main principles
- Operating systems – general knowledge
- Fluent use of revision control systems – git, svn

Publications

- A.Yanina, K.Vorontsov. Multimodal topic modeling for exploratory search in collective blog. *In Journal of Machine Learning and Data Analysis (to appear).*
- K.Vorontsov, O.Frei, A.Yanina, M.Apishev, P.Romov, M.Suvorova. Non-Bayessian Additive Regularization for Multimodal Topic Models on Large Collections. *In Proceedings of 24th ACM International Conference on Information and Knowledge Management (workshop Topic Models: Post-Processing and Applications)*, Melbourne, Australia, 2015
- K.Vorontsov, O.Frei, P.Romov, A.Yanina, M.Dudarenko, M.Apishev. BigARTM – open source library for topic modeling on big text collections. *In Proceedings of 17th International Conference "Data Analysis and Management in Data Intensive Domains" (DAMDID/RCDL)*, Obninsk, Russia, 2015

Programming projects

Python

- **Medical support system** (Feb 2015 to present).
  The service that uses machine learning algorithms for predicting correct treatment for patients and checking the medical prescriptions made by doctors. It may work as online-doctor and predict the disease from the symptoms. The system uses the data from clinical reports as a train sample. Also word embeddings are used.

- **Topic search (March 2015 to present)**
  A service for performing exploratory search through the articles. Topic models are built using latent Dirichlet allocation and additive regularization of topic models. A user is allowed to search using large texts or well-formulated queries.

- **Recommendation system of articles** (May 2014 – June 2015)
  Using topic model of articles from habrahabr collective blog (the most famous Russian IT-oriented blog), the service can recommend articles to users. It combines content-based approach with collaborative filtering and usage-based approach.

- **Simple text generator** (Oct 2014).
  By counting the three-grams in sample texts it can generate new novels using Markov chains.

- **Poetry authorship recognizer** (Nov 2014)**.**
  Naive Bayes classifier for recognizing the author of the poem. From already labeled data it can learn how to predict the author of an unlabeled poem.

- **Music genre classifier** (Nov 2014 - Dec 2014).
  Using mel frequency cepstral coeffitients as features it can determine the music genre of audio file.

C++

- **Parallel "clever" search of close documents, simple searching engine**
  (Feb 2014 – Jun 2014).
  The service for finding sets of relevant documents (close in meaning to one given text) within large text collections. The project also includes the realization of Kd-Tree for fast searching.

- **Parallel Priority queue based on a thread pool (March 2014).**
  It includes working with boost library: spinlock, shared_mutex, unique_lock, etc.

Github profile

- https://github.com/yanina-anastasia

Awards

- Participation in 11th International Conference on Intelligent Data Processing: Theory and Applications (presenting research "Multimodal Topic Models for Exploratory Search"), Barcelona, Oct 2016
- Participation in Moscow Data Science Meetup (presenting research "Open source library BigARTM for multimodal topic modeling"), Sep 2016
- Participation in 5th Symposium on Conformal and Probabilistic Prediction with Applications (presenting research "Exploratory Search Based on Topic Modeling"), Madrid, Apr 2016
- 3rd place in hackathon Skoltech Hack Race, Moscow, Feb 2016
- Participation in winter school and hackathon DeepHack.Q&A (The Allen AI Science Challenge), Moscow, Jan 2016
- Participation in Microsoft Machine Learning and Intelligence School, Saint-Petersburg, Jul 2015
- Taking part in Kaggle competitions: Caterpillar Tube Pricing, Liberty Mutual Group: Property Inspection Prediction, 2015
- Honorable Mention in the ACM ICPC, Moscow Sub regional of NEERC, 2012 – 2013
- Scholarship from the fund of development of innovative education, 2012-2016
- Honorable grant of the governor of the Samara region, 2012
- Russian National High School Mathematical Olympiad (region level) – 2nd prize, 2009-2010
- Russian National High School Physics Olympiad (region level) – 2nd prize, 2010-2011
- Russian National High School Informatics Olympiad (region level) – 4th prize, 2010-2011
- MSU High School Olympiad (mathematical section) – 3rd prize, 2012

Languages

- Russian - native
- English – fluent
  First Certificate in English, grade A (Level C1), **TOEFL ITP Test (Mar 2016)**
- French - beginner