Anastasiia Tsarkova

Education

September 2014 – June 2018 **BS in Applied Mathematics and Information Science**, *National Research University Higher School of Economics, Moscow, Russia*, GPA – 7.73 / 10 (transcript).

Projects

All projects are stored in git repository: `https://github.com/ilyshnikova/`

Work experience

July 2018 - now **Software Engineer**, *Google Switzerland*, Algorithmic Reconciliation for Knowledge Engine.

- Deprecated old Musical Recording Clustering pipeline and migrated it onto the horizontal infrastructure. Saved eng cost and maintenance cost (approx 1 engineer). Upon migration faced 70% quality drop, which I did an analysis for and came up with a proposal to bring the quality back to what it was. I also coordinated the whole migration work end to end, including finalizing rollout procedures.
- Migrated Google Podcasts Clustering System to the horizontal infrastructure. Quality analysis was required in order to guarantee quality parity. Interacted with 2 teams to achieve a consistent plan to finalize that migration.
- Worked on a new generation of Shopping Product Clustering pipeline, which was built from scratch. The pipeline processes billions of Shopping offers and catalogs, which required careful considerations with regards to resource usage.
- Became primary go-to person for Finance and Geo verticals of Knowledge Engine Reconciliation.
- Significantly improved freshness of Geo reconciliation from yearly to daily.
- Frequently debugging release blockers and becoming go-to person for analysis of such blockers.

September 2017 – March 2018 **Software Engineer**, *Yandex*, Advertising Services, Group of Product Development.
Continued to work in the team, which develops real-time advertisement service.

June 2017 – August 2017 **Business Intern**, *Google*, Trust & Safety, Gmail Abuse Group.
The main tasks of internship: to facilitate the infrastructure, automate the calculation of various statistics with map-reduce job, come up with metrics around precision and recall and create dashboards to monitor abuse.
Technologies: C++, Python, SQL

August 2016 – June 2017 **Software Engineer**, *Yandex*, Advertising Services, Real-time Technologies Development Group.
Working in a group developed high-load real-time service, which processes a large number of advertisement requests.
Responsibilities:

- Developing and maintaining product logic and infrastructural components
- Debugging and supporting the service

Tasks of real-time advertisement engine I worked on required algorithmic knowledge as well as good time management. Also, it was a useful experience of working in a rather large team and reading a lot of business-oriented code of a huge system.
Technologies: C++, Python, Perl, SQL, gdb

April 2016 – July 2016 **Software Engineer Intern**, *Yandex*, Geo-Informational Services, Real-time Services Group.
Implemented testing framework for the geo-service engine, which emulates loaded traffic on any given geographic graph. Took part in some of the infrastructural tasks of real-time search service.

Courseworks

2017–now **Improvement of Analytical DBMS ClickHouse, NRU HSE**, *NRU HSE, Diploma Project*.
ClickHouse is an open source column-oriented database management system capable of real-time generation of analytical data reports using SQL queries. My role for this project was to solve the following problems:

- To outline the possibility to force usage of ClickHouse's MergeTree engine table primary key in analyzing the query conditional, containing 'in'-clause with multiple fields. This feature was implemented and is already in the stable branch.
- To implement conditional computations for logical expressions such as 'and'/'or' operators and build a logic to identify in which cases conditional computations work faster. This part is in development state.

My pull requests can be found here

2016–2017 **Statistical Analysis of Financial Data by Means of Machine Learning,**, *NRU HSE, Team Project*.

- System of indicative statistical and qualitative analysis, which forecasts efficient investments
- Implemented library, which provides user with easy way to download and analyze financial data from different sources
- Includes human interface to facilitate work for users without programming skills

2015–2016 **Graphs Analyser Utility**, *NRU HSE*.

- Server-based program, which accepts time series points and detects anomalies on data series.
- Works online with complexity $\underline{O}(1)$ on point submission.
- Includes human html interface for controlling algorithms, which work on time series.
- Source code: `https://github.com/ilyshnikova/graph-analyzer`

Skills

Programming **C++, Python, Perl, JavaScript (basic), html, SQL**.

ML tools **Pandas, NumPy, SciPy, scikit-learn, matplotlib, Apache Spark, Keras, Vowpal Wabbit, word2vec**.

Other tools **linux, mac os, git, svn, deb packages, gdb, pdb, vim, flask, nginx**.

Languages **Russian,English**.