

- Experienced Software Engineer with a strong track record solving business problems using advanced algorithms, machine learning, and functional programming
- Proven ability to design and implement optimized data pipelines, perform data analysis, and deploy machine learning models in a production environment

Experience

- 01.2020–now **Senior Software Engineer, Dun & Bradstreet, San Jose, CA**
Company profiles, B2B. Data Engineering, Fuzzy search, Clustering
- Increased search hit rate by 4% by developing Spelling correction module. Applied Trie as BloomFilters, Count-min-sketch for language models. 7M unique words, 55M pairs, 55M triples.
 - Researched and developed a de-duplication method for 450M company profiles. Locality sensitive hashing, MinHash, Trie as BloomFilters for Hamming distance. Found 7M exact and fuzzy duplicates.
 - Implemented Verification and Identification services that match web domains to companies. 60M websites scale.
- Technologies: Scala, Python, Postgres, ML, AWS
- 07.2014–01.2020 **Senior Software Engineer, Piano Media, Izhevsk, Russia**
Platform for paywall and subscription business. Distributed systems, ML, Reports, DSL, Streaming. 200 Tb data in total, 5000 *rps_{avg}*, 8000 *rps_{peak}* daily.
- Launched Likely-to-Subscribe and Likely-to-Churn ML projects, 10M users
 - Led and built a project that provides various types of analytic reports, real-time, on-demand, customizable reports based on DSL. Connected Clojure-based DSL with Spark and BigQuery.
 - Implemented User Segmentation Engine, 100M users, used Bloom filter as KV-store
- Technologies: Scala, Clojure, Unix, Spark, HDFS, DSL, AWS, Akka, Kafka, Catboost, BigQuery
- 08.2013–07.2014 **Researcher, Senior Algorithm Engineer, Huawei Research, Moscow, Russia**
Nested data parallelism model, distributed systems, map-reduce, optimization compiler, meta-programming and code generation, domain specific languages
- Implemented Spark backend for nested data-parallelism model
 - Researched and developed efficient custom Spark RDD's for NDP-primitives
- Technologies: Scala, Spark, Distributed Programming, C++
- 12.2007–08.2013 **Senior Software Engineer, Rosneft, Izhevsk, Russia**
Oil mining, forecast, and monitoring system. OLAP, up to 2M data points per document, up to 15 dimensions, DSL, Calculations
- Increased calculation speed by 3.5 times using multithreading evaluations
 - Pioneered testing and verification via Partial evaluation
- Technologies: C#, Multithreading, DSL, Partial evaluations

Education

- 2010 **Ph.D., Udmurt State University, Institute of Control Sciences V.A. Trapeznikov Academy of Sciences, Izhevsk, Moscow, Russia**
Key words: online string algorithms, string matching, text indexing, dynamic suffix arrays, data compression. Advisor: prof. Nepejvoda N. N.
- 2006 **MS in Computer Science, Udmurt State University, Izhevsk, Russia**
Summa cum laude. Master thesis: Limited proof programming (Floyd-Hoare logic, annotated program, static verification, compiler)

CS Interests

graph theory, number theory, formal languages, string matching, data structures
domain-specific languages, functional programming

Skills

Data	Spark, Druid, Kafka, Rabbitmq, Big Query, Akka, SQL, Clickhouse	ML	LogReg, CatBoost, Transformers
Infrastructure	AWS, S3, EC2, LB, ECS, Docker	Culture	Tests, CI/CD, Code Reviews, Dashboards, Monitoring

ACM ICPC

coach 2012, World Final, Poland, Warsaw participant 2005, 2006, NEERC

Schools and conferences

DSLDI 2015 Summer School on DSL Design and Implementation, Aug 2015, Lausanne, Switzerland
EWSCS 2014 Estonian Winter School in Computer Science, Mar 2014, Palmse, Estonia
ALMADA 2013 School on Algorithms for Massive Data, Aug 2013, Moscow, Russia
Russir 2011 Russian Summer School in Information Retrieval, Aug 2011, Saint Petersburg, Russia
SSSEV 2011 Summer School in Software Engineering and Verification, Jul 2011, Moscow, Russia
Russir 2010 Russian Summer School in Information Retrieval, Sep 2010, Voronezh, Russia
conferences HighLoad++-2019, HighLoad++-2015, HighLoad++-2014, META-2016, META-2014, NSKF-2013, NSKF-2014, ULMCAMP-2014, Scalar-2017, STACHKA-2017, LambdaWorld-2017, f(by)-2017, Scalar-2018, Scalar-2019, f(by)-2019, f(by)-2020, fpure-2019, PyBay-2022

Main publications

D. V. Urbanovich, P. G. Ajtkulov, "Simple algorithm to maintain dynamic suffix array for text indexes", RuSSIR/EBDT, (2011), 40-45.
P. G. Ajtkulov, "Symbol array processing", UBS, 28 (2010), 126-178. (rus)

Talks

HighLoad-2019 Probabilistic data structures, <https://www.youtube.com/watch?v=Vx-l7KiIYuM>, <https://www.highload.ru/moscow/2019/abstracts/5944>
EWSCS 2014 19th Estonian Winter School in Computer Science, <http://cs.ioc.ee/ewscs/2014/index.php?page=../talks>
Online Suffix Array Construction

Collaboration

Dissernet <https://www.dissernet.org/> Plagiarism detection in Ph.D. theses and articles in Russia.
Algorithms on strings, ML/NLP, 1.2M theses and 2.6M articles, 500Gb raw texts

Misc

Translator Foundations for Programming Languages, John C. Mitchell. RCD press, 2010
Running Marathon: 3h22m, $\frac{1}{2}$ marathon: 1h38m, 10km: 44m30s
Chess Max Elo: 2118